

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:12:13 ; Search time 27.45 seconds  
(Without alignments)  
1153.300 Million cell updates/sec

Title: US-10-080-797-1  
Perfect score: 967  
Sequence: 1 HSHRDFPVLHVALNPLS.....SCHNAVIVLCIENSPMTASK 183

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTIST:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEA:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID     | Description         |
|------------|-------|-------------|--------|----|--------|---------------------|
| 1          | 840   | 86.9        | 184    | 11 | Q9JK63 | Q9JK63 mus musculus |
| 2          | 840   | 86.9        | 1140   | 11 | Q61434 | Q61434 mus musculus |
| 3          | 840   | 86.9        | 1774   | 11 | Q62001 | Q62001 mus musculus |
| 4          | 820   | 84.8        | 226    | 11 | Q9QZD2 | Q9QZD2 rattus norv  |
| 5          | 777   | 80.4        | 171    | 11 | Q9WUW5 | Q9WUW5 rattus norv  |
| 6          | 771   | 79.7        | 1344   | 13 | Q9J419 | Q9J419 gallus gall  |
| 7          | 722   | 74.7        | 160    | 11 | Q9CRT2 | Q9CRT2 mus musculus |
| 8          | 544   | 56.3        | 1367   | 11 | Q35206 | Q35206 mus musculus |
| 9          | 541   | 55.9        | 1367   | 11 | Q9E0D9 | Q9E0D9 mus musculus |
| 10         | 530   | 54.8        | 1388   | 4  | Q9Y4W4 | Q9Y4W4 homo sapien  |
| 11         | 529   | 54.7        | 102    | 4  | Q96T70 | Q96T70 homo sapien  |
| 12         | 430.5 | 44.5        | 581    | 5  | Q9V509 | Q9V509 drosophila   |
| 13         | 373   | 38.6        | 650    | 5  | Q17866 | Q17866 caenorhabdi  |
| 14         | 373   | 38.6        | 778    | 5  | Q9U9K6 | Q9U9K6 caenorhabdi  |
| 15         | 373   | 38.6        | 864    | 5  | Q93336 | Q93336 caenorhabdi  |
| 16         | 373   | 38.6        | 1117   | 5  | Q9U9K7 | Q9U9K7 caenorhabdi  |

|    |      |     |      |    |        |                     |
|----|------|-----|------|----|--------|---------------------|
| 17 | 88.5 | 9.2 | 498  | 2  | Q9KXK2 | Q9KXK2 streptomyce  |
| 18 | 84   | 8.7 | 493  | 16 | Q98AT9 | Q98AT9 rhizobium 1  |
| 19 | 83.5 | 8.6 | 208  | 16 | Q92K28 | Q92K28 rhizobium m  |
| 20 | 80   | 8.3 | 651  | 5  | Q9VFA9 | Q9VFA9 drosophila   |
| 21 | 79.5 | 8.2 | 1715 | 6  | Q9GLM4 | Q9GLM4 bos taurus   |
| 22 | 79   | 8.2 | 477  | 10 | Q9SMY7 | Q9SMY7 arabidopsis  |
| 23 | 79   | 8.2 | 525  | 10 | Q94JL8 | Q94JL8 arabidopsis  |
| 24 | 78.5 | 8.1 | 285  | 16 | Q98E05 | Q98E05 rhizobium 1  |
| 25 | 78   | 8.1 | 539  | 2  | Q9L4Y1 | Q9L4Y1 streptomyce  |
| 26 | 78   | 8.1 | 564  | 2  | Q93XN8 | Q93XN8 streptomyce  |
| 27 | 77.5 | 8.0 | 395  | 11 | Q900P4 | Q900P4 mus musculus |
| 28 | 77   | 8.0 | 1056 | 4  | Q9UDT9 | Q9UDT9 homo sapien  |
| 29 | 76.5 | 7.9 | 421  | 16 | Q98NL4 | Q98NL4 rhizobium 1  |
| 30 | 76   | 7.9 | 204  | 4  | Q96G49 | Q96G49 homo sapien  |
| 31 | 76   | 7.9 | 636  | 4  | Q90624 | Q90624 homo sapien  |
| 32 | 75.5 | 7.8 | 334  | 11 | Q9MTJ8 | Q9MTJ8 mus musculus |
| 33 | 75.5 | 7.8 | 346  | 5  | Q9VNS3 | Q9VNS3 drosophila   |
| 34 | 75.5 | 7.8 | 904  | 10 | Q48541 | Q48541 hordeum vul  |
| 35 | 75   | 7.8 | 351  | 16 | Q99X31 | Q99X31 staphylococ  |
| 36 | 74.5 | 7.7 | 850  | 4  | Q14425 | Q14425 homo sapien  |
| 37 | 74.5 | 7.7 | 962  | 10 | Q957S8 | Q957S8 hordeum vul  |
| 38 | 74   | 7.7 | 919  | 15 | Q9WPP0 | Q9WPP0 chimpanzee   |
| 39 | 74   | 7.7 | 960  | 2  | Q9KTY0 | Q9KTY0 streptomyce  |
| 40 | 74   | 7.7 | 1063 | 4  | Q9NSW6 | Q9NSW6 homo sapien  |
| 41 | 74   | 7.7 | 1063 | 4  | Q96CY0 | Q96CY0 homo sapien  |
| 42 | 74   | 7.7 | 1066 | 4  | Q9NZS3 | Q9NZS3 homo sapien  |
| 43 | 74   | 7.7 | 1215 | 4  | Q94975 | Q94975 homo sapien  |
| 44 | 73.5 | 7.6 | 325  | 4  | Q9Y247 | Q9Y247 homo sapien  |
| 45 | 73.5 | 7.6 | 905  | 10 | Q9FYI0 | Q9FYI0 hordeum vul  |

## ALIGNMENTS

| RESULT | ID   | Q9JK63                                    | PRELIMINARY: | PRT: | 184 AA. |
|--------|--|---|--------------|------|---------|
| AC     | Q9JK63   | Q9JK63                                    |              |      |         |
| DT     | 01-OCF-2000  | (TREMUREL. 15, Created)                   |              |      |         |
| DT     | 01-OCF-2000  | (TREMUREL. 15, Last sequence update)      |              |      |         |
| DT     | 01-DEC-2001  | (TREMUREL. 19, Last annotation update)    |              |      |         |
| DE     | ENDOSTATIN (FRAGMENT)  |   |              |      |         |
| OS     | Mus musculus (Mouse)   |   |              |      |         |
| OC     | Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:  |   |              |      |         |
| OC     | Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Mus.          |   |              |      |         |
| NC     | NCBI_TaxID=10090;  |   |              |      |         |
| RN     | [1]  |   |              |      |         |
| RP     | SEQUENCE FROM N.A.   |   |              |      |         |
| RC     | STRAIN=CHINESE KUNMING;  |   |              |      |         |
| RA     | Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;         |   |              |      |         |
| RT     | "Anticancer treatment of targeted fusion protein delivery to tumor |   |              |      |         |
| RT     | neovascularization";   |   |              |      |         |
| RL     | Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.            |   |              |      |         |
| DR     | EMBL: AF257775; AAF69009.1; -                                      |   |              |      |         |
| DR     | HSSP: P39061; IKOE.  |   |              |      |         |
| FT     | NON_TER  | 1   |              |      |         |
| FT     | NON_TER  | 184                                       |              |      |         |
| FT     | SEQUENCE   | 184 AA; 20376 MW; AC06F9DBD103412A CRC64; |              |      |         |

Query Match 86.9%; Score 840; DB 11; Length 184;  
Best Local Similarity 85.6%; Pred. No. 2.4e-78;  
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

|    |     |  |     |
|----|-----|--|-----|
| QY | 1   | HSHRDFPVLHVALNPLSGMGRGICADPFCQARAVGLACTFRAPLSSRLQDIYSI     | 60  |
| DB | 1   | HTHDDFPVLHVALNPLSGMGRGICADPFCQARAVGLCTFRAPLSSRLQDIYSI      | 60  |
| QY | 61  | VRRDRAAVPVLNKKDLPLSPWEALFSGSGPLKPGARIFSPFGKVLRRHPMPKQSW    | 120 |
| DB | 61  | VRRDRAAVPVLNKKDLPLSPWEALFSGSGPLKPGARIFSPFGKVLRRHPMPKQSW    | 120 |
| QY | 121 | HGSDPGRRLTESYCTWRTREAPSATGQASSLLGRLIGQASASCHNAVIVLCIENSPMT | 180 |

DB 121 HSDSPGRLMESYCECTWRTETTGATGOASLSLGRLLLEQKASCHNSYIVLCIENSFMT 180  
 OY 181 A 181  
 DB 181 S 181

## RESULT 2

061434 PRELIMINARY: PRT: 1140 AA.

AC 061434:  
 DT 01-NOV-1996 (TREMBLREL. 01, Created)  
 DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLREL. 19, Last annotation update)  
 DE COLLAGEN (FRAGMENT).  
 GN COL15A1  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;  
 RT "Identification of a novel collagen chain represented by extensive  
 RT interruptions in the triple-helical region."  
 RL Cell. Mol. Biol. Res. 196:576-582(1993).  
 DR EMBL: D17546; BAA04483.1; -.  
 DR HSSP: P39061; IKOE.  
 DR MGD: MGI:88449; COL15a1.  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 6.  
 FT NON\_TER 1  
 SQ SEQUENCE 1140 AA; 115156 MW; 8B0C7E6862B3BDPE CRC64;

Query Match 86.9%; Score 840; DB 11; Length 1140;  
 Best Local Similarity 85.6%; Pred. No. 2,6e-77;  
 Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

OY 1 HSHRDFQPVHLHVALNPSGSGMRGIRGADFOCFQOARAVGLAGTFRATLSRLDLYSI 60  
 DB 957 HTHQDFQPVHLHVALNTPSLSGSGMRGIRGADFOCFQOARAVGLSGTFRATLSRLDLYSI 1016  
 OY 61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVW 120  
 DB 1017 VRRADGSPVIVNLKDEVLSFSGSQGLQPGARIFSPDGKDVLRHPTWPKSVW 1076  
 OY 121 HSDPGRRLTESYCECTWRTETTGATGOASLSLGRLLLEQKASCHNSYIVLCIENSFMT 180  
 DB 1077 HSDSPGRLMESYCECTWRTETTGATGOASLSLGRLLLEQKASCHNSYIVLCIENSFMT 1136  
 OY 181 A 181  
 DB 1137 S 1137

RESULT 3  
 062001 PRELIMINARY: PRT: 1774 AA.

AC 062001:  
 DT 01-NOV-1996 (TREMBLREL. 01, Created)  
 DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)  
 DT 01-OCT-2001 (TREMBLREL. 18, Last annotation update)  
 DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 PRECURSOR (XVIII) COLLAGEN  
 DE (PROCOLLAGEN, TYPE XVIII, ALPHA 1) (ALPHA-1 TYPE XVIII COLLAGEN).  
 GN COL18A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [11]  
 RP SEQUENCE FROM N.A.

RC TISSUE-PRIMARY TAIL CULTURE;  
 RX MEDLINE=94245707; PubMed=8188673;  
 RA Rehn M., Hentikka E., Philafantemi T.;  
 RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,  
 RT partial structure of the corresponding gene, and comparison of the  
 RT alpha 1(xviii) chain with its homologue, the alpha 1(xv) collagen  
 RT chain."  
 RL J. Biol. Chem. 269:13929-13935(1994).  
 RN [2]  
 RP SEQUENCE OF 1-562 FROM N.A.  
 RX MEDLINE=94240112; PubMed=8183894;  
 RA Rehn M., Philafantemi T.;  
 RT "Alpha 1(xviii), a collagen chain with frequent interruptions in the  
 RT collagenous sequence, a distinct tissue distribution, and homology  
 RT with type xv collagen."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).  
 RN [3]

RP SEQUENCE OF 1-562 FROM N.A.  
 RX MEDLINE=95181468; PubMed=7876242;  
 RA Rehn M., Philafantemi T.;  
 RT "Identification of three N-terminal ends of type XVIII collagen chains  
 RT and tissue-specific differences in the expression of the corresponding  
 RT transcripts. The longest form contains a novel motif homologous to rat  
 RT and Drosophila frizzled proteins."  
 RL J. Biol. Chem. 270:4705-4711(1995).  
 DR EMBL: U03715; AAC52903.1; JOINED.  
 DR EMBL: U03716; AAC52903.1; JOINED.  
 DR EMBL: U03718; AAC52903.1; JOINED.  
 DR EMBL: U34607; AAC52903.1; JOINED.  
 DR EMBL: U34608; AAC52903.1; JOINED.  
 DR EMBL: U34609; AAC52903.1; JOINED.  
 DR EMBL: U34610; AAC52903.1; JOINED.  
 DR EMBL: U34611; AAC52903.1; JOINED.  
 DR EMBL: U34612; AAC52903.1; JOINED.  
 DR EMBL: U34613; AAC52903.1; JOINED.  
 DR EMBL: U11637; AAC52179.1; -.  
 DR HSSP: P39061; IKOE.  
 DR MGD: MGI:88451; COL18a1.  
 DR InterPro: IPR000087; Collagen.  
 DR InterPro: IPR000024; Fz\_domain.  
 DR InterPro: IPR001791; Lamlnln\_G.  
 DR InterPro: IPR003129; TSPN.  
 DR Pfam: PF01392; Fz; 1.  
 DR Pfam: PF01391; Collagen; 6.  
 DR Pfam: PF01392; Fz; 1.  
 DR SMART: SM00210; TSPN; 1.  
 DR SMART: SM00603; FRI; 1.  
 DR SMART: SM00282; LamG; 1.  
 DR SMART: SM00210; TSPN; 1.  
 DR PROSITE: PS50038; Fz; 1.  
 KW Signal.  
 SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 86.9%; Score 840; DB 11; Length 1774;  
 Best Local Similarity 85.6%; Pred. No. 4,7e-77;  
 Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;

OY 1 HSHRDFQPVHLHVALNPSGSGMRGIRGADFOCFQOARAVGLAGTFRATLSRLDLYSI 60  
 DB 1591 HTHQDFQPVHLHVALNTPSLSGSGMRGIRGADFOCFQOARAVGLSGTFRATLSRLDLYSI 1650  
 OY 61 VRRADRAAVPIVNLKDELLFPSWEALFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSVW 120  
 DB 1651 VRRADGSPVIVNLKDEVLSFSGSQGLQPGARIFSPDGKDVLRHPTWPKSVW 1710  
 OY 121 HSDPGRRLTESYCECTWRTETTGATGOASLSLGRLLLEQKASCHNSYIVLCIENSFMT 180  
 DB 1711 HSDSPGRLMESYCECTWRTETTGATGOASLSLGRLLLEQKASCHNSYIVLCIENSFMT 1770  
 OY 181 A 181  
 DB 1771 S 1771

```

RESULT 4
ID 090ZD2 PRELIMINARY; PRT: 226 AA.
AC 090ZD2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE COLLAGEN XVIII (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=20227226; PubMed=10766159;
RA Perletti G., Concarl P., Giardin R., Marras E., Piccinini F.,
RA Folkmann J., Chen L.;
RT "Antitumor activity of endostatin against carcinogen-induced rat
RT primary mammary tumors."
RL Cancer Res. 60:1793-1796(2000).
DR EMBL; AF189709; AAF00975.1; -.
DR HSSP; P39061; IKOE.
FT NON_TER
SQ SEQUENCE 226 AA; 25350 MW; 38B83C0486CE949 CRC64;

```

```

Query Match 84.8%; Score 820; DB 11; Length 226;
Best Local Similarity 84.0%; Pred. No. 3.5e-76;
Matches 152; Conservative 14; Mismatches 15; Indels 0; Gaps 0;

```

```

QY 1 HSHRDPOVHLVALNSPLSGMRGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSI 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 43 HTHQDHPVHLVALNPLPSGMRGIRGADFOCFQOARAVGLSTFRFAFLSSRLQDLYSI 102
QY 61 VRRADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 103 VRRADSSVPIVNLKDELFPSSWDTLFSSGQQLHSGARIFSPDGRVLRHPTWPKSVW 162
QY 121 HGSDDPGRRLTESYCTWRTEAPSATGQASSLLGRLGQSAASHAYIVLCIENSFMT 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 163 HGSDDPGRRLTESYCTWRTEATGVTGQASSLLGRLLEOKAESCHNSYIVLCIENSFMT 222
QY 181 A 181
DB 223 S 223

```

```

RESULT 5
ID 09WUW5 PRELIMINARY; PRT: 171 AA.
AC 09WUW5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE COLLAGEN TYPE XVIII, ALPHA (I) CHAIN (FRAGMENT).
GN COL18A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;
RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Jia J.D., Bauer M., Sedlacek N., Ruehl M., Riecken E.O., Schuppan D.;
RT "Temporopartial expression of collagen XVIII/endostatin in acute and

```

```

RT chronic liver injuries.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ236873; CAB44263.1; -.
DR HSSP; P39061; IKOE.
FT NON_TER
FT NON_TER
SQ SEQUENCE 171 AA; 18933 MW; 81BE2EE3FC2C8E72 CRC64;

```

```

Query Match 80.4%; Score 777; DB 11; Length 171;
Best Local Similarity 85.4%; Pred. No. 6.3e-72;
Matches 146; Conservative 11; Mismatches 14; Indels 0; Gaps 0;

```

```

QY 9 VLHLVALNSPLSGMRGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSI VRRADRAA 68
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 VLHLVALNPLPSGMRGIRGADFOCFQOARAVGLSTFRFAFLSSRLQDLYSI VRRADRSS 60
QY 69 VPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW HGSDDPGR 128
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 VPIVNLKDELFPSSWDTLFSSGQQLHSGARIFSPDGRVLRHPTWPKSVW HGSDDPGR 120
QY 129 RLTESYCTWRTEAPSATGQASSLLGRLGQSAASHAYIVLCIENSFMT 179
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 121 RLMSYCTWRTEATGVTGQASSLLSGRLLEOKAESCHNSYIVLCIENSFMT 171

```

```

RESULT 6
ID 093419 PRELIMINARY; PRT: 1344 AA.
AC 093419;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DE COLLAGEN XVIII PRECURSOR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411346; PubMed=9738008;
RA Halfter W., Dong S., Schurer B., Cole G.J.;
RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan.";
RL J. Biol. Chem. 273:25404-25412(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Halfter W., Dong S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083440; AAC33294.2; -.
DR HSSP; P39061; IKOE.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen_7.
DR Pfam; PF02210; TSPN_1.
DR SMART; SM00282; TSPN; 1.
DR SMART; SM00210; TSPN; 1.
KW Signal.
FT SIGNAL
SQ SEQUENCE 1344 AA; 137402 MW; 7AA366E4FE940CCD CRC64;

```

```

Query Match 79.7%; Score 771; DB 13; Length 1344;
Best Local Similarity 77.6%; Pred. No. 3.9e-70;
Matches 142; Conservative 18; Mismatches 23; Indels 0; Gaps 0;

```

```

QY 1 HSHRDPOVHLVALNSPLSGMRGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSI 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1161 HVHDDPQALHLVALNPLPSGMRGIRGADFOCFQOARAVGLAGTFRAFLSSRLQDLYSI 1220
QY 61 VRRADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 1221 VRRADTFAPVIVLKDVEVLSNMEALFTGSEAPLRAGARILSFGDHDLDQDSAMPQKSIW 1280
Oy 1221 HGSDFNGRRLTESYCEWTREAPSATGQASLLGRLGQASACHHAYIVLCIENSFMT 180
    ||||| ||||| ||||| : ||||| : ||||| ||||| ||||| |||||
Db 1281 HGSDFNGRRLTESYCEWTREAPSATGQASLLGRLGQASACHHAYIVLCIENSFMT 1340
Oy 181 ASK 183
    |||
Db 1341 AAK 1343

RESULT 7
OQCRT2 PRELIMINARY: PRT: 160 AA.
ID OQCRT2:
AC OQCRT2:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROCOLLAGEN, TYPE XVIII, ALPHA 1 (FRAGMENT).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H. A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schiml L. M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boftelli D., Bojunga N., Carninci P., de Bonaldo M. F.,
RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D. A., Kamuya M., Lee N. H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K. -F.,
RA Suzuki H., Toyo-oka K., Wang K. H., Weltz C., Whitaker C., Wilming L.,
RA Wyszynski B. Y., Yoshida K., Hasegawa Y., Kawai H., Kohlsuk S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK014292; BAB29249.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88451; Col18a1.
FT NON_TER 1
SQ SEQUENCE 160 AA; 1725 MW; 60F853D77C375D2 CRC64;

```

```

Query Match 74.7%; Score 722; DB 11; Length 160;
Best Local Similarity 85.4%; Pred. No. 2, 6e-66;
Matches 134; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

```

```

Oy 25 GTRGADFOGFOQARAVGLAGTFRAFSLSRLODLYSVRRADAAVIVNKLDELPPSME 84
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GTRGADFOGFOQARAVGLAGTFRAFSLSRLODLYSVRRADAAVIVNKLDELPPSMD 60
Oy 85 ALFSGSEGLKPGARIFSDGKDVLRHPTWPKSVHSGSDPNRRLTESYCEWTREAPS 144
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 61 SLFSSGQGLQGCARIFSDGKDVLRHPTWPKSVHSGSDPNRRLTESYCEWTREAPS 120
Oy 145 ATGQASSLLGRLGQASACHHAYIVLCIENSFMTA 181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 ATGQASSLLGRLGQASACHHAYIVLCIENSFMTS 157
RESULT 8

```

```

O35206
ID O35206 PRELIMINARY: PRT: 1367 AA.
AC O35206:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97480713; PubMed=9339358;
RA Hag P. M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT corresponding gene to 4B1-3. Comparison of mouse and human alpha 1
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains.";
RL Genomics 45:31-41(1997).
RX EMBL: AF011450; AAC53387.1; -.
DR HSSP: P39061; IKOE.
DR MGD: MGI:88449; Col15a1.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001791; Laminin_G.
DR InterPro: IPR003129; TSPN.
DR Pfam: PF02210; Collagen; 4.
DR Pfam: PF02210; TSPN; 1.
DR SMART: SM00282; LamG; 1.
DR SMART: SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1367 AA; 140525 MW; A483A1254AF3AEBC CRC64;

```

```

Query Match 56.3%; Score 544; DB 11; Length 1367;
Best Local Similarity 58.2%; Pred. No. 7, 9e-47;
Matches 103; Conservative 29; Mismatches 41; Indels 4; Gaps 1;

```

```

Oy 7 OPLHLVIALNSPLSGMRGIRGADFOGFOQARAVGLAGTFRAFSLSRLODLYSVRRADR 66
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 1194 RPLVHLVIALNTPVAGDIR---ADFQFOQARAGLSTFRFLSHLDLSTVVKAKR 1249
Oy 67 AAVPIVNLKDELFPSEWALFSGSEGLKPGARIFSDGKDVLRHPTWPKSVHSGSDPN 126
    : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||
Db 1250 FGLPIVNLKGVLFENMWSIFSGDGQFMTHPIYFSGDRDWTDPSPQKVVYHGSNNH 1309
Oy 127 GRLTESYCEWTREAPSATGQASLLGRLGQASACHHAYIVLCIENSFMTASK 183
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1310 GVLVDKCYCEAWRTDMAVTGFAFSLSTGKILDKAKVSCANRLIVLCIENSFMTDTR 1366

RESULT 9
O3EOD9 PRELIMINARY: PRT: 1367 AA.
ID O3EOD9:
AC O3EOD9:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TYPE XV COLLAGEN.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97480713; PubMed=9339358;
RA Hag P. M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the

```



DB 1331 GVRVDNYCEAMRTADTAVTGLASPLSTGKILDQKAYSCANRLIVLCIENSFMTDAR 1387

RESULT 11

ID 096770 PRELIMINARY: PRT: 102 AA.

AC 096770: 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI\_TaxID=9606;

RP SEQUENCE FROM N.A.

RA Delinger M.H., Trautmann K., Schluesener H.J.;

RT "Endostatin Promotes delayed secondary damage following traumatic brain injury."

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF333247; AAK50626.1; -

FT NON\_TER 1 1

FT SEQUENCE 102 AA; 11147 MW; ECAC47AA6420947D CRC64;

Query Match 54.7%; Score 529; DB 4; Length 102;  
Best Local Similarity 98.0%; Pred. No. 9.1e-47;  
Matches 100; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 71 IYNLDELFPSEWALFSGSEGLPKGARIFFSDGKDLRHPTWPKSVWHSQDPNGRR 130  
DB 1 IYNLKKELLPSEWALFSGSEGLPKGARIFFSDGKDLRHPTWPKSVWHSQDPNGRR 60

OY 131 TESYCTWRTPEAPSATGQASSLLGGRLLGQSAASCHAVIYL 172  
DB 61 TESYCTWRTPEAPSATGQASSLLGGRLLGQSAASCHAVIYL 102

RESULT 12

ID 09VS09 PRELIMINARY: PRT: 581 AA.

AC 09VS09: 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE CG8645 PROTEIN.

GN CG8645.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OC NCBI\_TaxID=7227;

RP SEQUENCE FROM N.A.

RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Abmayyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borovica D., Botchan M.R., Bouck J., Brokstein P., Brotlier P., Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de la Piedad B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Doudouk K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durkin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Houtin D., Houston K.A., Howland T.J., Wei M.-H., Ibegyan C., Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPreston D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pachet J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spler E., Spradling A.C., Stapleton M., Strong R., Sun E., Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."

RL Science 287:2185-2195(2000).

DR EMBL: AE003560; AAF50621.1; -

DR HSSP: P39061; IKOE.

DR FlyBase: FBgn0035732; CG8645.

DR InterPro: IPR000087; Collagen.

DR InterPro: IPR002088; PPTA.

DR Pfam: PF01391; Collagen\_3.

DR PROSITE: PS00904; PPTA; UNKNOWN\_1.

SQ SEQUENCE 581 AA; 60772 MW; 19EC1E4BCB477FE7 CRC64;

Query Match 44.5%; Score 430.5; DB 5; Length 581;  
Best Local Similarity 48.8%; Pred. No. 1.2e-35;  
Matches 82; Conservative 29; Mismatches 56; Indels 1; Gaps 1;

OY 8 PYLHLVALNSPLSGMGIRGADFCGQARAVGLAGTERAFELSLRLDLYSIVRRADRA 67  
DB 373 PYLHLVALNSPLSGMGIRGADFCGQARAVGLAGTERAFELSLRLDLYSIVRRADRA 431

OY 68 AYPVNLKDELFPSEWALFSGSEGLPKGARIFFSDGKDLRHPTWPKSVWHSQDPNG 127  
DB 432 AYPVNLKDELFPSEWALFSGSEGLPKGARIFFSDGKDLRHPTWPKSVWHSQDPNG 491

OY 128 RLTESYCTWRTPEAPSATGQASSLLGGRLLGQSAASCHAVIYLCTE 175  
DB 492 RLTESYCTWRTPEAPSATGQASSLLGGRLLGQSAASCHAVIYLCTE 539

RESULT 13

ID 017866 PRELIMINARY: PRT: 650 AA.

AC 017866: 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE CLE-1C PROTEIN.

GN CLE-1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea; Rhabdilitidae; Pelodertinae; Caenorhabditis.

OC NCBI\_TaxID=6239;

RP SEQUENCE FROM N.A.

RA MEDLINE=21157401; PubMed=11257122;

RA Ackley B.D., Crew J.R., Elamha H., Pihlajantem T., Kuo C.J., Kramer J.M.;

RT "The NC1/endostatin domain of Caenorhabditis elegans type XVIII collagen affects cell migration and axon guidance."

RL J. Cell Biol. 152:1219-1232(2001).

DR EMBL: AF164959; AAD47825.1; -

DR HSSP: P39061; IKOE.

DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen; 2.  
SQ SEQUENCE 650 AA; 69596 MW; 6CF29ED9C16B170E CRC64;

Query Match 38.6%; Score 373; DB 5; Length 650;  
Best Local Similarity 44.1%; Pred. No. 1.3e-29;  
Matches 75; Conservative 26; Mismatches 63; Indels 6; Gaps 4;

QY 9 VLHLVNLSPGSGMGRGADFCFOQARAVGLAGTFRAFLSRLODLXSYRRADRAA 68  
DB 464 VIMHIALSQPFGSGNLHGLRGADLCYREARAAGTTTFRAMLSNVODLVRIHVSVD-FD 522  
QY 69 VPIVNLKDELFPSEWALFSGSEGPKLPGARIFSFDDGKDVLRHPTWPKSVHSGSDPNGR 128  
DB 523 TTVVNVAGHLLFPSSWRSFVNGAQ--MNPBKLFSFDRHVDLNDLRWPDKRVHSGSKDGI 580  
QY 129 RLTESYCETWTEAPSATGQASSLGGRLGQSAAS--CHHAYIVLCIEN 176  
DB 581 R-ABOYCDGWRRADSLTSLAGHISNTSIFQSSGSEKCNKLVLCVEN 629

RESULT 14  
Q909K6 PRELIMINARY; PRT; 778 AA.

AC Q909K6;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE CLE-1B PROTEIN.  
GN CLE-1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditidae.  
OX NCBI\_TaxID=6239;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21157401; PubMed=11257122;  
RA Ackley B.D., Crew J.R., Elamaa H., Pihlajaniemi T., Kuo C.J.,  
RA Kramer J.M.;  
RT "The NCL/endostatin domain of Caenorhabditis elegans type XVIII  
RT collagen affects cell migration and axon guidance."  
RL J. Cell Biol. 152:1219-1232(2001).  
DR EMBL: AF164959; AAD47824.1; -  
DR HSSP: P39061; IKOE;  
DR InterPro: IPR000087; Collagen.  
DR InterPro: IPR003129; TSPN.  
DR Pfam: PF01391; Collagen; 2.  
DR SMART: SM00210; TSPN; 1.  
SQ SEQUENCE 778 AA; 83787 MW; 743A7FC8A7FF1ABD CRC64;

Query Match 38.6%; Score 373; DB 5; Length 778;  
Best Local Similarity 44.1%; Pred. No. 1.3e-29;  
Matches 75; Conservative 26; Mismatches 63; Indels 6; Gaps 4;

QY 9 VLHLVNLSPGSGMGRGADFCFOQARAVGLAGTFRAFLSRLODLXSYRRADRAA 68  
DB 592 VIMHIALSQPFGSGNLHGLRGADLCYREARAAGTTTFRAMLSNVODLVRIHVSVD-FD 650  
QY 69 VPIVNLKDELFPSEWALFSGSEGPKLPGARIFSFDDGKDVLRHPTWPKSVHSGSDPNGR 128  
DB 651 TTVVNVAGHLLFPSSWRSFVNGAQ--MNPBKLFSFDRHVDLNDLRWPDKRVHSGSKDGI 708  
QY 129 RLTESYCETWTEAPSATGQASSLGGRLGQSAAS--CHHAYIVLCIEN 176  
DB 709 R-ABOYCDGWRRADSLTSLAGHISNTSIFQSSGSEKCNKLVLCVEN 757

RESULT 15  
Q93336 PRELIMINARY; PRT; 864 AA.  
ID Q93336  
AC Q93336;

DT 01-FEB-1997 (TREMblrel. 02, Created)  
DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
DE C36B1.1 PROTEIN.  
GN C36B1.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditidae.  
OX NCBI\_TaxID=6239;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Lennard N.;  
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C. elegans: A platform for  
RT investigating biology."  
RL Science 282:2012-2018(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA White S.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Z80215; CAB02278.2; -  
DR EMBL: Z81079; CAB02278.2; JOINED.  
DR EMBL: Z81079; CAC42298.1; -  
DR EMBL: Z80215; CAC42298.1; JOINED.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen; 2.  
SQ SEQUENCE 864 AA; 93341 MW; 4A8EDB9B960E56F CRC64;

Query Match 38.6%; Score 373; DB 5; Length 864;  
Best Local Similarity 44.1%; Pred. No. 1.3e-29;  
Matches 75; Conservative 26; Mismatches 63; Indels 6; Gaps 4;

QY 9 VLHLVNLSPGSGMGRGADFCFOQARAVGLAGTFRAFLSRLODLXSYRRADRAA 68  
DB 678 VIMHIALSQPFGSGNLHGLRGADLCYREARAAGTTTFRAMLSNVODLVRIHVSVD-FD 736  
QY 69 VPIVNLKDELFPSEWALFSGSEGPKLPGARIFSFDDGKDVLRHPTWPKSVHSGSDPNGR 128  
DB 737 TTVVNVAGHLLFPSSWRSFVNGAQ--MNPBKLFSFDRHVDLNDLRWPDKRVHSGSKDGI 794  
QY 129 RLTESYCETWTEAPSATGQASSLGGRLGQSAAS--CHHAYIVLCIEN 176  
DB 795 R-ABOYCDGWRRADSLTSLAGHISNTSIFQSSGSEKCNKLVLCVEN 843

Search completed: August 12, 2002, 10:14:25  
Job time: 132 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:12:33 ; Search time 11.91 Seconds

(without alignments)  
594,935 Million cell updates/sec

Title: US-10-080-797-1

Perfect score: 967  
Sequence: 1 HSHRDFQVHLVALNSPLS.....SCHHAYIVLCIENSEFMTASK 183

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

| Result No. | Score | Query | Match Length | ID         | Description        |
|------------|-------|-------|--------------|------------|--------------------|
| 1          | 967   | 100.0 | 1516         | CA1H_HUMAN | P39060 homo sapien |
| 2          | 840   | 86.9  | 1527         | CA1H_MOUSE | P39061 mus musculu |
| 3          | 529   | 54.7  | 1388         | CA1E_HUMAN | P39059 homo sapien |
| 4          | 76    | 7.9   | 556          | GLI_CHICK  | P55878 gallus gall |
| 5          | 75    | 7.8   | 271          | Y4BG_RHISN | P55374 rhizobium s |
| 6          | 74    | 7.7   | 296          | CRK_XENTLA | P87378 xenopus lae |
| 7          | 74    | 7.7   | 780          | STRN_RAT   | P70483 rattus norv |
| 8          | 74    | 7.7   | 1215         | HD46_HUMAN | Q94077 homo sapien |
| 9          | 73    | 7.5   | 512          | PPX_ECOLI  | P29014 escherichia |
| 10         | 72.5  | 7.5   | 884          | YP67_MYCTU | Q50654 mycobacteri |
| 11         | 72    | 7.4   | 304          | CRK_HUMAN  | P46108 homo sapien |
| 12         | 72    | 7.4   | 304          | CRK_MOUSE  | O64010 mus musculu |
| 13         | 72    | 7.4   | 304          | CRK_RAT    | O63768 rattus norv |
| 14         | 72    | 7.4   | 613          | HS75_CCNAL | P87222 candida alb |
| 15         | 70.5  | 7.3   | 312          | NR1L_CHICK | P55806 gallus gall |
| 16         | 70    | 7.2   | 780          | STRN_MOUSE | O55106 mus musculu |
| 17         | 69    | 7.1   | 266          | CB21_SINAI | P13851 sinapis alb |
| 18         | 69    | 7.1   | 309          | YHCC_ECOLI | P45476 escherichia |
| 19         | 69    | 7.1   | 326          | THOE_PSEME | O00460 pseudomonas |
| 20         | 69    | 7.1   | 364          | YM28_MYCTU | Q10512 mycobacteri |
| 21         | 69    | 7.1   | 462          | MYCN_MOUSE | P39066 mus musculu |
| 22         | 69    | 7.1   | 1233         | NME3_HUMAN | Q14957 homo sapien |
| 23         | 68.5  | 7.1   | 390          | PGK_BUCAI  | P57525 buchera ap  |
| 24         | 68.5  | 7.1   | 1289         | C5AB_BACUD | O45753 bacillus th |
| 25         | 68    | 7.0   | 853          | CAR4_HUMAN | Q9Y239 homo sapien |
| 26         | 68    | 7.0   | 999          | MERK_HUMAN | Q12866 homo sapien |
| 27         | 68    | 7.0   | 3491         | ERY1_SACER | O03131 saccharopol |
| 28         | 67.5  | 7.0   | 145          | ANF_RANCA  | P18909 rana catesb |
| 29         | 67.5  | 7.0   | 317          | GGH_RAT    | O62867 rattus norv |
| 30         | 67.5  | 7.0   | 536          | CAR9_HUMAN | O9H257 homo sapien |
| 31         | 67.5  | 7.0   | 579          | FAAH_MOUSE | O08914 mus musculu |
| 32         | 67.5  | 7.0   | 594          | MG11_MOUSE | O60710 mus musculu |
| 33         | 67.5  | 7.0   | 628          | PRPE_ECOLI | P77495 escherichia |

|    |      |     |      |   |            |                    |
|----|------|-----|------|---|------------|--------------------|
| 34 | 67.5 | 7.0 | 692  | 1 | GYRB_BARBA | P94281 bartonella  |
| 35 | 67.5 | 7.0 | 1367 | 1 | VGR2_MOUSE | P35918 mus musculu |
| 36 | 67.5 | 7.0 | 3206 | 1 | POLG_PSBMV | P29152 p genome po |
| 37 | 67   | 6.9 | 250  | 1 | HXB9_HUMAN | P17482 homo sapien |
| 38 | 67   | 6.9 | 577  | 1 | BAG3_MOUSE | Q91471 mus musculu |
| 39 | 67   | 6.9 | 956  | 1 | RRPO_SBMV  | P21405 southern be |
| 40 | 66.5 | 6.9 | 368  | 1 | GALT_HUMAN | O60795 homo sapien |
| 41 | 66.5 | 6.9 | 449  | 1 | DHE2_CLOSY | P24295 clostridium |
| 42 | 66.5 | 6.9 | 683  | 1 | ACSI_KILUA | O60011 kluyveromyc |
| 43 | 66.5 | 6.9 | 1548 | 1 | SMCY_MOUSE | O62240 mus musculu |
| 44 | 66   | 6.8 | 250  | 1 | HXB9_MOUSE | P20615 mus musculu |
| 45 | 66   | 6.8 | 295  | 1 | MURB_RICCN | Q921t8 ticketisia  |

## ALIGNMENTS

RESULT 1  
ID CA1H\_HUMAN STANDARD: PRT; 1516 AA.  
AC P39060: 09Y608: 09Y607: 09UK38;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Collagen alpha 1(XVII) chain precursor [contains: Endostatin].  
GN COL18A1.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxId=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98164096; PubMed=9503365;  
RA Szarela J., Ylikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;  
RT Complete primary structure of two variant forms of human type XVIII  
RT collagen and tissue-specific differences in the expression of the  
RT corresponding transcripts.";  
RL Matrix Biol. 16:319-328(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20289799; PubMed=10830953;  
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,  
RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,  
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald A., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
RA Rosenthal K., Kudon J., Shibuya K., Kawasaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,  
RA Schafie M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H.,  
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,  
RA Weinmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
RA Leinrach H., Reinhardt R., Yaspo M.-L.;  
RT "The DNA sequence of human chromosome 21.";  
RL Nature 405:311-319(2000).  
RN [3]  
RP SEQUENCE OF 834-1516 FROM N.A.  
RX MEDLINE=94245237; PubMed=8188291;  
RA Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,  
RA Olsen B.R.;  
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen  
RT and localization of the alpha 1(XVII) collagen gene to mouse  
RT chromosome 10 and human chromosome 21.";  
RL Genomics 19:494-499(1994).  
RN [4]  
RP SEQUENCE OF 1334-1516 FROM N.A.  
RX TISSUE=Placenta;  
RA Zhi-Fong H., Biao L., Wei-Jie Z., Xiang-Fu W.;  
RT "Cloning and expression of human endostatin gene in Escherichia  
RT coli.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP INVOLVEMENT IN KNOBLOCH SYNDROME.

RA MEDLINE-20400145; PubMed-10942434;  
RA Sertie A.L., Sossel V., Camargo A.A., Zatz M., Brahe C.,  
RA Passos-Bueno M.R.;  
RT "Collagen XVII, containing an endogenous inhibitor of angiogenesis  
RT and tumor growth, plays a critical role in the maintenance of retinal  
RT structure and in neural tube closure.";  
RL Hum. Mol. Genet. 9:2051-2058(2000).  
RN [6]  
RP VARIANT ASN-1437.  
RX PubMed-11606364;  
RA Iugnetti P., Suzuki O., Godol P.H., Alves V.A., Sertie A.L.,  
RA Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,  
RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;  
RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes  
RT for the development of prostatic adenocarcinoma.";  
RL Cancer Res. 61:7375-7378(2001).  
CC -1- FUNCTION: COLA18A PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE  
CC RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.  
CC -1- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL  
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
CC BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH  
CC FACTOR SIGNALING.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A SHORT FORM/NC1-303 AND A LONG  
CC FORM/NC-483 (SHOWN HERE). ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS  
CC IN LIVER, LUNG AND KIDNEY.  
CC -1- PTM: POLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -1- POLYMORPHISM: There is an association between a polymorphism in  
CC position 1437 and prostate cancer. Heterozygous Asn-1437  
CC individuals have a 2.5 times increased chance of developing  
CC prostate cancer as compared with homozygous Asp-1437 individuals.  
CC -1- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KS  
CC or KNO): an autosomal recessive disorder defined by the occurrence  
CC of high myopia, vitreoretinal degeneration with retinal  
CC detachment, macular abnormalities and occipital encephalocele.  
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
CC INTERRUPTED HELICES (FACIT) FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF018081: AAC39658.1; -;  
DR EMBL: AF018082: AAC39659.1; -;  
DR EMBL: AL163302: CAB90482.1; -;  
DR EMBL: L22548: AAS1864.1; -;  
DR EMBL: AF184060: AAF01310.1; ALT\_INIT.  
DR HSSP: P39061: 1KOE.  
DR GLYCONSULEDB: P39060: -.  
DR MIM: 120328: -.  
DR MIM: 267730: -.  
DR InterPro: IPR000087: Collagen.  
DR InterPro: IPR001791: Laminin\_G.  
DR InterPro: IPR003129: TSPN.  
DR Pfam: PF01391: Collagen; 6.  
DR Pfam: PF02210: TSPN; 1.  
DR SMART: SM00282; LamG; 1.  
DR SMART: SM00210; TSPN; 1.  
DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;  
KW Polymorphism.  
FT SIGNAL 1 23 POTENTIAL.  
FT CHAIN 24 1516 COLLAGEN ALPHA 1(XVII) CHAIN.  
FT CHAIN 1334 1516 ENDOSTATIN.  
FT DOMAIN 24 516 NONHELICAL REGION 1 (NC1).  
FT DOMAIN 517 550 TRIPLE-HELICAL REGION 1 (COL1).  
FT DOMAIN 551 560 NONHELICAL REGION 2 (NC2).  
FT DOMAIN 561 640 TRIPLE-HELICAL REGION 2 (COL2).

FT DOMAIN 641 664 NONHELICAL REGION 3 (NC3).  
FT DOMAIN 665 786 TRIPLE-HELICAL REGION 3 (COL3).  
FT DOMAIN 787 809 NONHELICAL REGION 4 (NC4).  
FT DOMAIN 810 892 TRIPLE-HELICAL REGION 4 (COL4).  
FT DOMAIN 893 906 NONHELICAL REGION 5 (NC5).  
FT DOMAIN 907 948 TRIPLE-HELICAL REGION 5 (COL5).  
FT DOMAIN 949 961 TRIPLE-HELICAL REGION 6 (NC6).  
FT DOMAIN 962 1034 TRIPLE-HELICAL REGION 6 (COL6).  
FT DOMAIN 1035 1044 NONHELICAL REGION 7 (NC7).  
FT DOMAIN 1045 1077 TRIPLE-HELICAL REGION 7 (COL7).  
FT DOMAIN 1078 1089 NONHELICAL REGION 8 (NC8).  
FT DOMAIN 1090 1111 TRIPLE-HELICAL REGION 8 (COL8).  
FT DOMAIN 1112 1118 NONHELICAL REGION 9 (NC9).  
FT DOMAIN 1119 1173 TRIPLE-HELICAL REGION 9 (COL9).  
FT DOMAIN 1174 1186 NONHELICAL REGION 10 (NC10).  
FT DOMAIN 1187 1204 TRIPLE-HELICAL REGION 10 (COL10).  
FT DOMAIN 1205 1516 NONHELICAL REGION 11 (NC11).  
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 164 164 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 691 691 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1329 1329 O-LINKED (GALNAC. . .).  
FT DISULFID 1366 1506 BY SIMILARITY.  
FT DISULFID 1468 1498 CELL ATTACHMENT SITE (POTENTIAL).  
FT SITE 1095 1097 MISSING (IN SHORT ISOFORM).  
FT VARSPIC 181 215 HTTEAGTLPAPRPSGPRAPLGSVPSPSS -> MA  
FT PRCPPMPRRRLDLVLAIPVLLGVRAASAE (IN  
FT SHORT ISOFORM).  
FT VARIANT 1437 1437 D -> N.  
FT CONFLICT 428 428 /FTID=VAR\_012709.  
FT CONFLICT 841 841 F -> S (IN REF. 2).  
FT CONFLICT 841 841 I -> V (IN REF. 2).  
FT CONFLICT 877 877 V -> L (IN REF. 3).  
FT CONFLICT 886 886 P -> R (IN REF. 3).  
FT CONFLICT 912 912 P -> R (IN REF. 3).  
FT CONFLICT 933 933 R -> L (IN REF. 3).  
FT CONFLICT 975 975 P -> L (IN REF. 3).  
FT CONFLICT 1064 1064 A -> P (IN REF. 3).  
FT CONFLICT 1084 1084 L -> K (IN REF. 3).  
FT CONFLICT 1120 1120 P -> A (IN REF. 3).  
FT CONFLICT 1123 1123 P -> A (IN REF. 3).  
FT CONFLICT 1126 1126 P -> PPGP (IN REF. 2).  
FT CONFLICT 1206 1206 G -> GO (IN REF. 3).  
FT CONFLICT 1304 1304 R -> G (IN REF. 3).  
FT CONFLICT 1314 1314 A -> G (IN REF. 3).  
FT CONFLICT 1324 1324 IR -> CG (IN REF. 3).  
FT CONFLICT 1443 1443 R -> T (IN REF. 4).  
FT CONFLICT 1483 1483 S -> Y (IN REF. 4).  
SQ SEQUENCE 1516 AA; 153840 MW; 3C70F29A4476EB76 CRC64;

Query Match 100.0%; Score 967; DB 1; Length 1516;  
Best Local Similarity 100.0%; Pred. No. 2; Be-86;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HSHRDFQPLVHLVALNPSGSGMRGIRGADFCFOQARAVAGTGFRAFLSLRLDIVSI 60  
DB 1334 HSHRDFQPLVHLVALNPSGSGMRGIRGADFCFOQARAVAGTGFRAFLSLRLDIVSI 1393  
OY 61 VRRADRAAVPIVNLKDELLFPSEWALEFSGSEGPPLRGARIFSFDEKDVLRHPTWPOKSYW 120  
DB 1394 VRRADRAAVPIVNLKDELLFPSEWALEFSGSEGPPLRGARIFSFDEKDVLRHPTWPOKSYW 1453  
OY 121 HGSDPNGRRLTESYCTWTWTAPSATGQASSLLGRLGLQSSAASCHHATIVYCIENSFWT 180  
DB 1454 HGSDPNGRRLTESYCTWTWTAPSATGQASSLLGRLGLQSSAASCHHATIVYCIENSFWT 1513  
OY 181 ASK 183  
DB 1514 ASK 1516

```

RESULT 2
CAH MOUSE
ID CAH MOUSE STANDARD; PRT; 1527 AA.
AC P39061; 062002; Q61437;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(XVIII) chain precursor [contains: Endostatin].
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC STRAIN=BALE/C; TISSUE=Liver;
RX MEDLINE=94245707; PubMed=818673;
RA Rehn M.V., Hintikka E., Pihlajaniemi T.;
RT Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain.;
RN [2]
RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RA Rehn M., Hintikka E., Pihlajaniemi T.;
RT Characterization of the mouse gene for the alpha-1 chain of type
RT XVIII collagen (COL18A1) reveals that the three variant N-terminal
RT polypeptide forms are transcribed from two widely separated
RT promoters.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).
RX MEDLINE=94240112; PubMed=8183894;
RA Rehn M.V., Pihlajaniemi T.;
RT Alpha 1(XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.;
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [4]
RP SEQUENCE OF 240-1527 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=94240111; PubMed=8183893;
RA Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;
RT Isolation and sequencing of cDNAs for proteins with multiple domains
RT of Gly-xaa-yaa repeats identify a distinct family of collagenous
RT proteins.;
RL Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).
RN [5]
RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE=97160848; PubMed=9008168;
RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vassios G., Lane W.S.,
RA Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;
RT Endostatin: an endogenous inhibitor of angiogenesis and tumor
RT growth.;
RL Cell 88:277-285(1997).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.
RX MEDLINE=98169382; PubMed=9501087;
RA Hohenester E., Sasaki T., Olsen B.R., Timpl R.;
RT Crystal structure of the angiogenesis inhibitor endostatin at 1.5-A
RT resolution.;
RL EMBO J. 17:1656-1664(1998).
CC -1- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL
CC PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY
CC BINDING TO THE HEPARAN SULFATE PROTEOGLYCANS INVOLVED IN GROWTH
CC FACTOR SIGNALLING.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH

```

```

CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L16898; AAA37434.1; -
CC EMBL; U03714; AAA20657.1; -
CC EMBL; U03715; AAC52901.1; -
CC EMBL; U34606; AAC52901.1; JOINED.
CC EMBL; U34608; AAC52901.1; JOINED.
CC EMBL; U34609; AAC52901.1; JOINED.
CC EMBL; U34610; AAC52901.1; JOINED.
CC EMBL; U34611; AAC52901.1; JOINED.
CC EMBL; U34612; AAC52901.1; JOINED.
CC EMBL; U34613; AAC52901.1; JOINED.
CC EMBL; U03716; AAC52901.1; JOINED.
CC EMBL; U03718; AAC52901.1; JOINED.
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC52902.1; JOINED.
CC EMBL; L22545; AAA18787.1; -
CC EMBL; U03715; AAC52902.1; -
CC EMBL; U34607; AAC52902.1; JOINED.
CC EMBL; U34608; AAC52902.1; JOINED.
CC EMBL; U34609; AAC52902.1; JOINED.
CC EMBL; U34610; AAC52902.1; JOINED.
CC EMBL; U34611; AAC52902.1; JOINED.
CC EMBL; U34612; AAC52902.1; JOINED.
CC EMBL; U34613; AAC52902.1; JOINED.
CC EMBL; U03716; AAC52902.1; JOINED.
CC EMBL; U03718; AAC5290
```





DB 260 -KOEIARAETWK 271

RESULT 6

CRK\_XENLA STANDARD; PRT: 296 AA.

ID CRK\_XENLA

AC P87378;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE SH2/SH3 adaptor crk (Adaptor molecule crk) (CRK).

GN CRK.

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;

OC Xenopodinae; Xenopus.

OX NCBI\_TaxID=8335;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE-97180926; PubMed-9029144;

RA Evans E.K., Lu W., Strum S.L., Mayer B.J., Kornbluth S.;

RT "Crk is required for apoptosis in Xenopus egg extracts.";

RL EMBO J. 16:230-241(1997)

CC -1- FUNCTION: REQUIRED FOR APOPTOSIS IN XENOPUS EGG EXTRACTS.

CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.

CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: U89774; ABA49698.1; -

DR HSP: 064010; 1CKA.

DR InterPro: IPR000980; SH2.

DR InterPro: IPR001452; SH3.

DR Pfam: PF00017; SH2; 1.

DR Pfam: PF00018; SH3; 1.

DR PRINTS: PR00401; SH2DOMAIN.

DR PRINTS: PR00452; SH3DOMAIN.

DR SMART: SM00252; SH2; 1.

DR SMART: SM00326; SH3; 2.

DR PROSITE: PS50001; SH2; 1.

DR PROSITE: PS50002; SH3; 1.

KW SH2 domain; SH3 domain; Repeat.

FT DOMAIN 13 112 SH2.

FT DOMAIN 125 185 SH3 1.

FT DOMAIN 249 289 SH3 2.

FT SEQUENCE 296 AA; 33409 MW; 544F11F4A1F75A66 CRC64;

Query Match 7.7%; Score 74; DB 1; Length 296;

Best Local Similarity 32.4%; Pred. No. 5.4;

Matches 24; Conservative 12; Mismatches 20; Indels 18; Gaps 5;

OY 87 FGSSEG---PLKPGARIFSGDKVLRHPTWPKSWHGSDPGR--LTSEYCEWRTPE 141

DB 136 FIGNDEDEDLPFKK-----DIIRIKREKPEQWMAEDSDGRGMIPVIVEXR-- 184

OY 142 APSATGQASSLLGG 155

DB 185 PSSPG--SALIGG 196

RESULT 7

STRN\_RAT STANDARD; PRT: 780 AA.

ID STRN\_RAT

AC P70483;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Striatin.

GN STRN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 222-243; 258-269; 451-456; 488-494

RP AND 764-777.

RC STRAIN-WISTAR; TISSUE=Brain;

RX MEDLINE-96354919; PubMed-8769426;

RA Castets F., Bartoli M., Barnier J.V., Baillet G., Salin P.,

RA Moglich A., Bourgois J.-P., Denizot F., Rougon G., Calothy G.,

RA Monnerot A.;

RT "A novel calmodulin-binding protein, belonging to the WD-repeat

RT family, is localized in dendrites of a subset of CNS neurons.";

RL J. Cell Biol. 134:1051-1062(1996).

CC -1- FUNCTION: BINDS CALMODULIN IN A CALCIUM DEPENDENT MANNER. MAY

CC FUNCTION AS SCAFFOLDING OR SIGNALING PROTEIN.

CC -1- SUBUNIT: INTERACTS WITH PROTEIN PHOSPHATASE 2A (PP2A) (POTENTIAL).

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC AND MEMBRANE-BOUND.

CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN THE CENTRAL NERVOUS

CC SYSTEM. MOSTLY CONFINED IN DENDRITES, NOT IN AXONS, AND IS MOST

CC ABUNDANT IN DENDRITIC SPINES.

CC -1- SIMILARITY: BELONGS TO THE STRIATIN FAMILY OF WD-REPEAT PROTEINS.

CC -1- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL: X99326; CAA67701.1; -

DR InterPro: IPR001680; WD40.

DR Pfam: PF00400; WD40; 7.

DR PRINTS: PR00320; GPROTEINBPT.

DR SMART: SM00320; WD40; 5.

DR PROSITE: PS00678; WD\_REPEATS\_1; 2.

DR PROSITE: PS50082; WD\_REPEATS\_2; 4.

KW Calmodulin-binding; Repeat; WD repeat; Coiled coil.

FT DOMAIN 53 120 COILED COIL (POTENTIAL).

FT DOMAIN 149 166 CALMODULIN-BINDING (POTENTIAL).

FT REPEAT 461 500 WD 1.

FT REPEAT 514 553 WD 2.

FT REPEAT 567 606 WD 3.

FT REPEAT 662 701 WD 4.

FT REPEAT 704 743 WD 5.

FT REPEAT 750 779 WD 6.

FT SITE 55 63 CAVEOLIN-BINDING (POTENTIAL).

FT SEQUENCE 780 AA; 86226 MW; 67F7C2209D560F8 CRC64;

Query Match 7.7%; Score 74; DB 1; Length 780;

Best Local Similarity 23.7%; Pred. No. 17;

Matches 51; Conservative 19; Mismatches 65; Indels 80; Gaps 12;

OY 14 ALNSPLSGMRCIRGADFOCARAVG--LACTFRAFLSSRLQDLYSIVRADRAVPIV 72

DB 398 ALTFPPSSGSKFTMGAD--EALSELGLGLAG-----LTVANEADSLAYDIA 443

OY 73 NUKDELLFPSWEALFSGSGEPLKPGARIFSPDGKDLVR--HPTW----- 115

DB 444 NKKDALR--KTMWPKFT-----LRS-----HFDGIRALAPHPLEPVLITASEDHLLKMMNL 492

OY 116 -----QKSVWHSQSDP-----NCRRL-----TESYCEWRTPEASATG 147

DB 493 QKTAAPAKKSTSLDVEPIYTFRAHKGFLCVWSSNCEQCTSGGTDLGDSSTINPNVDP 552

```

OY 148 QAS---SLGRLGOS-----AASCHAYILC 173
DB 553 YDSYDPSVLRGPIRGHTDAVWGLAYSAAHQRLISC 587

RESULT 8
HDA6_HUMAN
ID HDA6_HUMAN STANDARD: PRT; 1215 AA.
AC 09UN7:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Histone deacetylase 6 (HD6).
GN HDAC6 OR JM21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
(1)
RN SEQUENCE FROM N.A.
RX MEDLINE=99238449; PubMed=10220385;
RA Grozinger C.M., Hassig C.A., Schreiber S.L.;
RT "Three proteins define a class of human histone deacetylases related
to yeast Hda1p."
RL Proc. Natl. Acad. Sci. U.S.A. 96:4868-4873(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strom T.M., Gutwilling N., Nakamura G., Hellbrand H., Drescher B.,
RA Rosenthal A., Meindl A.;
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: RESPONSIBLE FOR THE DEACETYLATION OF LYSINE RESIDUES ON
CC THE N-TERMINAL PART OF THE CORE HISTONES (H2A, H2B, H3 AND H4).
CC HISTONE DEACETYLATION PLAYS AN IMPORTANT ROLE IN TRANSCRIPTIONAL
CC REGULATION, CELL CYCLE PROGRESSION AND DEVELOPMENTAL EVENTS (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: BELONGS TO THE HISTONE DEACETYLASE / ACUC / APHA
CC FAMILY. HD SUBFAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF132609; AAD29048.1; -
DR EMBL: AF011972; CAA09893.1; -
DR MIM: 300272; -
DR InterPro: IPR000286; His_deacetylase.
DR InterPro: IPR001607; zf-UBP.
DR Pfam: PF00850; Hist_deacetyl; 2.
DR Pfam: PF02148; zf-UBP; 1.
DR PRINTS: PR01270; HDASUPER.
DR SMART: SM00290; znf-UBP; 1.
KW Hydrolyase; Nuclear protein; Repeat.
FT DOMAIN 87 404 HISTONE DEACETYLASE 1.
FT DOMAIN 482 800 HISTONE DEACETYLASE 2.
SQ SEQUENCE 1215 AA; 131430 MW; 77406633FB11CFA CRC64;

Query Match 7.7%; Score 74; DB 1; Length 1215;
Best Local Similarity 23.3%; Pred. No. 28;
Matches 52; Conservative 18; Mismatches 83; Indels 70; Gaps 10;

OY 5 DDPVPLHLVNLN-SPLSGMRGIRGADFCQFQGAR-AVGLAG-----TFRAFL 50
DB 333 EFPQPLVLAAGDALGDPKGMATPAGFAUTLHLMGLAGSKLLSLGGYINERALA 392
OY 51 SSRLODLYSIVRRAD-----RAAVPIVNLKDELLFSPWEALF----- 87

```

```

DB 393 EGVASLHTLL--GDPCPMLESPGAPCRSAQASVSCALEALEPEWVLYVSTERVDNM 450
OY 88 -----SGSEGLPKPARIFSPGKDYLRHPMP--QKSWHGSDNGRRLLSYCETV 138
DB 451 EEDNVESEEGEPWEP-----PVLPLTWVPLQSRGTGLVYDON---MNNHCNLM 496
OY 139 RTEAPSATGQASSL-----LGRLGOSAAASCHAYILC 173
DB 497 DSHHEVFORLIRIMCRLEGLAGRLCTLTPRPATEHELLTC 539

RESULT 9
ID PXX_ECOLI STANDARD: PRT; 512 AA.
AC P29014; P76981;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Exopolysphatase (EC 3.6.1.11) (Exopolypase) (Metaphosphatase).
GN PXX OR B2502 OR Z3765 OR ECS3364.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
(1)
RN SEQUENCE FROM N.A. AND SEQUENCE OF 1-10.
RP STRAIN=K12;
RC MEDLINE=93107072; PubMed=8380170;
RA Akiyama M., Crooke E., Kornberg A.;
RT "An exopolysphatase of Escherichia coli. The enzyme and its ppx
RT gene in a polysphatase operon."
RL J. Biol. Chem. 268:633-639(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / M61655;
RX MEDLINE=97426617; PubMed=92278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Colado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsunashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakanura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubdaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Gotlock E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;

```

RX MEDLINE-21156231; PubMed-11258796;  
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kohara S., Shiba T., Hattori M., Shindagawa H.;  
RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
RL DNA Res. 8:111-22(2001).  
RN (6)  
RP SIMILARITY TO GPPA.  
RA MEDLINE-94025037; PubMed-8212131;  
RX Reizer J., Reizer A., Saler M.H. Jr., Bork B., Sander C.;  
RT "Exopolyphosphate phosphatase and guanosine pentaphosphate  
RT phosphatase belong to the sugar kinase/actin/hsp 70 superfamily.";  
RL Trends Biochem. Sci. 18:247-248(1993).  
CC -1- FUNCTION: DEGRADATION OF INORGANIC POLYPHOSPHATES. ORTHOPHOSPHATE  
CC IS RELEASED PROGRESSIVELY FROM THE ENDS OF POLYPHOSPHATE OF CIRCA  
CC 500 RESIDUES LONG, BUT CHAINS OF CIRCA 15 RESIDUES COMPLETE POORLY  
CC WITH POLYPHOSPHATE AS SUBSTRATE.  
CC -1- CATALYTIC ACTIVITY: [polyposphate](n) + H(2)O =  
CC [polyphosphate](n-1) + phosphate.  
CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Membrane-associated.  
CC -1- SIMILARITY: BELONGS TO THE GPPA / PPX FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: L06129; AAA24415.1; -  
DR EMBL: AE000336; AAC75555.1; -  
DR EMBL: D90878; BAA16390.1; -  
DR EMBL: D90880; BAA16392.1; -  
DR EMBL: AE005479; AAG57612.1; -  
DR EMBL: AP002561; BAB36787.1; -  
DR PIR: A45333; A45333.  
DR Ecogene: EGI1403; ppx.  
DR InterPro: IPR003695; ppx.GPPA.  
DR Pfam: PF02541; Ppx-Gppa; 1.  
KW Hydrolyase; Magnesium; Membrane; Complete proteome.  
FT INIT\_MER 0  
FT SEQUENCE 512 AA; 58004 MW; 48611AEF5D9FB9C3 CRC64;  
-----  
Query Match 7.5%; Score 73; DB 1; Length 512;  
Best Local Similarity 21.9%; Pred. No. 13;  
Matches 46; Conservative 28; Mismatches 66; Indels 70; Gaps 11;  
-----  
OY 34 PGOAR-----AVLAGTFERA-----FLSSRLDLY 58  
DB 186 FORARMAAOKLETTWOPRIGGWNVAMGASGITKAHEVLYMGEKDGITTPERLEKLY 245  
OY 59 -STVRADRAAVPIVNLKDE--LLPFSWEALFSG-----SEGPLKPGARIFSP 103  
DB 246 KEVLRRNFASLSTPLSEBRKTVFVPGGLALICGVFDALARELRISDGLARGV-LYEM 304  
OY 104 DGR-----DVLRRPTWPKQSWMGSDPNGRRL--TESYCTWTTEAPS-ATGQASSLLG- 154  
DB 305 EGRFRHODVRSRTASSLANOYHIDSEQARRVLDITQMYEOWMEQREOPKLANHPOLEALLRW 364  
OY 155 -----GRLLGSAASCHHAYIVLCIENS 177  
DB 365 AAMLEHVGILNINHSGLHRHSAYI-----LONS 391  
-----  
RESULT 10  
YP67\_MYCTU STANDARD; PRT; 884 AA.  
-----

AC 050654; Q50731.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 95.4 kDa protein RV2567.  
GN RV2567 OR MT2643 OR MTCY227.34C OR MTCY964.01C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxId=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-H37RV;  
RX MEDLINE-98295967; PubMed-9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeller K., Gas S., Barry C.E. III, Tekela F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emdinova M.D., Salzberg S.L.,  
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: SOME, TO M.TUBERCULOSIS RV2411C AND STNECHOCYSTIS PCC  
CC 6803 SLD0335.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: Z77250; CAB01053.1; -  
DR EMBL: AE007098; AAK46956.1; -  
DR TIGR: MT2643; -  
DR Tuberculist; RV2567; -  
KW Hypothetical protein; Complete proteome.  
FT CONFLICT 645 645 0 -> R (IN REF. 2).  
FT SEQUENCE 884 AA; 95448 MW; 95D23A4D2DEB365 CRC64;  
-----  
Query Match 7.5%; Score 72.5; DB 1; Length 884;  
Best Local Similarity 29.5%; Pred. No. 27;  
Matches 31; Conservative 15; Mismatches 26; Indels 33; Gaps 6;  
-----  
OY 63 RADRAVPIVNLKDELLPFSWEALFSGSEGPLKPGARIFSPG-----KDVLRHPW 114  
DB 609 RADMIIVA-----PS--TLWSILYVDPDPGSLVOSVEELALAAQAVRDOLSDNHW 656  
OY 115 -----PKQSWMGSDPNGRRLTESYCTWTTEAPSATGQASSLLG 154  
DB 657 MYLANVERAVERHKSDP-----FQSLAE---ADAVLASAQAETLAG 693  
-----  
RESULT 11  
CRK\_HUMAN STANDARD; PRT; 304 AA.  
AC P46108;  
-----





DR PDB: 1CKB; 08-MAY-95.  
DR PDB: 1B07; 06-JAN-99.  
DR MGD: MGI:88508; Crko.  
DR InterPro: IPR000980; SH2.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00017; SH2.1.  
DR Pfam: PF00018; SH3.1.  
DR PRINTS: PR00401; SH2DOMAIN.  
DR PRINTS: PR00452; SH3DOMAIN.  
DR SMART: SM00252; SH2.1.  
DR SMART: SM00326; SH3.2.  
DR PROSITE: PSS0001; SH2.1.  
DR PROSITE: PSS0002; SH3.1.  
KW Proto-oncogene; SH2 domain; SH3 domain; Repeat; Alternative splicing;  
KW Phosphorylation; 3D-structure.  
FT DOMAIN 13 118 SH2.  
FT DOMAIN 132 192 SH3.1.  
FT DOMAIN 256 286 SH3.2.  
FT VARSPLIC 205 304 MISSING (IN ISOFORM CRK-I).  
SQ SEQUENCE 304 AA; 33814 MW; 5491896FC7A89065 CRC64;

Query Match 7.4%; Score 72; DB 1; Length 304;  
Best Local Similarity 27.8%; Pred. No. 8.8;  
Matches 20; Conservative 12; Mismatches 24; Indels 16; Gaps 3;

OY 100 IFSFDDK-----DVLRHPTWPKSVHSGSDPNGR--LTESYCEWTREASATG 147  
DB 140 LFDPNGNDEDLFPKKGDILIRDKPEEOWMNAEDSEGRKGMIPVYEYK----PASA 195

OY 148 QASLLGRLG 159  
DB 196 SVSALLGNQEG 207

RESULT 13

CRK\_RAT STANDARD; PRT; 304 AA.

AC Q63768;  
DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Proto-oncogene C-crk (P38) (Adapter molecule crk).

GN CRK.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI\_TaxID=10116;

RN (1)  
RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;  
RA MEDLINE=97057214; PubMed=8901553;

RA Kizaka-Kondoh S., Matsuda M., Okayama H.;  
RT "Crkl signals from epidermal growth factor receptor to Ras.";

RL Proc. Natl. Acad. Sci. U.S.A. 93:12177-12182(1996).

CC -1- FUNCTION: THE CRK-I AND CRK-II FORMS DIFFER IN THEIR BIOLOGICAL  
CC ACTIVITIES. CRK-II HAS LESS TRANSFORMING ACTIVITY THAN CRK-I. BOTH  
CC CRK-I AND CRK-II BIND TO MANY OF TYROSINE-PHOSPHORYLATED PROTEINS  
CC THAT BIND TO GRB2.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: CRK-I AND CRK-II (SHOWN HERE);  
CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE CRK-I FORM LACKS THE  
CC LAST 100 RESIDUES.

CC -1- TISSUE SPECIFICITY: CRK-II IS EXPRESSED IN ALL TISSUES AND CELLS  
CC WHEREAS CRK-I IS EXPRESSED AT LOWER LEVEL AND IN LIMITED CELL-  
CC TYPES.

CC -1- DOMAIN: THE C-TERMINAL SH3 DOMAIN FUNCTION AS A NEGATIVE MODULATOR  
CC FOR TRANSFORMATION AND THE N-TERMINAL SH3 DOMAIN APPEARS TO  
CC FUNCTION AS A POSITIVE REGULATOR FOR TRANSFORMATION (BY  
CC SIMILARITY).

CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.

CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: D4481; BAA07924.1; -  
DR HSP: O64010; 1CKA.  
DR InterPro: IPR000980; SH2.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00017; SH2.1.  
DR Pfam: PF00018; SH3.1.  
DR PRINTS: PR00401; SH2DOMAIN.  
DR PRINTS: PR00452; SH3DOMAIN.  
DR SMART: SM00252; SH2.1.  
DR SMART: SM00326; SH3.2.  
DR PROSITE: PSS0001; SH2.1.  
DR PROSITE: PSS0002; SH3.1.  
KW Proto-oncogene; SH2 domain; SH3 domain; Repeat; Alternative splicing;  
KW Phosphorylation.  
FT DOMAIN 13 118 SH2.  
FT DOMAIN 132 192 SH3.1.  
FT DOMAIN 256 286 SH3.2.  
FT VARSPLIC 205 304 MISSING (IN ISOFORM CRK-I).  
FT VARIANT 244 244 Q -> R (NRK-23 INACTIVE MUTANT).  
FT VARIANT 253 253 K -> E (NRK-23 INACTIVE MUTANT).  
SQ SEQUENCE 304 AA; 33844 MW; 4CFBF65BE72E265 CRC64;

Query Match 7.4%; Score 72; DB 1; Length 304;  
Best Local Similarity 27.8%; Pred. No. 8.8;  
Matches 20; Conservative 12; Mismatches 24; Indels 16; Gaps 3;

OY 100 IFSFDDK-----DVLRHPTWPKSVHSGSDPNGR--LTESYCEWTREASATG 147  
DB 140 LFDPNGNDEDLFPKKGDILIRDKPEEOWMNAEDSEGRKGMIPVYEYK----PASA 195

OY 148 QASLLGRLG 159  
DB 196 SVSALLGNQEG 207

RESULT 14

HS75.CANAL STANDARD; PRT; 613 AA.

AC P87232;  
DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Heat shock protein SSB1.

GN SSB1 OR HSP70B.

OS Candida albicans (Yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Microsporid Saccharomycetales; Candida.

OX NCBI\_TaxID=5476;

RN (1)  
RP SEQUENCE FROM N.A.

RC STRAIN=MO-1;  
RA MEDLINE=97344370; PubMed=9200817;

RA Manen V., Cervera A.M., Martinez J.P., Gozalbo D.;  
RT "Molecular cloning of a Candida albicans gene (SSB1) coding for a  
RT protein related to the Hsp70 family.";

RL Yeast 13:677-681(1997).

CC -1- FUNCTION: MAY AID IN THE PASSAGE OF THE NASCENT POLYPEPTIDE CHAIN  
CC THROUGH THE RIBOSOME CHANNEL INTO THE CYTOSOL. SUCH AN INTERACTION  
CC COULD BE CRUCIAL FOR CONTINUOUS TRANSPORT OF THE POLYPEPTIDE;  
CC COULD SERVE TO PREVENT THE NASCENT POLYPEPTIDE FROM INTERFERING  
CC WITH TRANSLATION BY CLOGGING THE RIBOSOME CHANNEL (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH TRANSLATING  
CC RIBOSOMES; MAY BIND DIRECTLY TO THE NASCENT POLYPEPTIDE (BY  
CC SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 70 FAMILY.

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X97723; CAA66308.1; -.
DR HSSP: P08107; IHUO.
DR COMPLEYEAST-2DPAGE; P87222; -.
DR InterPro: IPR001023; HSP70.
DR Pfam: PF00012; HSP70; 1.
DR PRINTS: PR00301; HEATSHOCK70.
DR PROSITE: PS00297; HSP70_1; 1.
DR PROSITE: PS00329; HSP70_2; 1.
DR PROSITE: PS01036; HSP70_3; 1.
DR Heat shock; ATP-binding; Multigene family; Protein biosynthesis.
KW SEQUENCE 613 AA; 66432 MW; 5FBA9F8F99327F9 CRC64;

Query Match 7.4%; Score 72; DB 1; Length 613;
Best Local Similarity 25.6%; Pred. No. 20;
Matches 33; Conservative 20; Mismatches 50; Indels 26; Gaps 6;

QY 44 GTFRAFLSSRLDLYSIVRADRAAVPIVNLKDELFPSEWALFS-----GSGGPL 94
DB 4 GVFQGAIGIDLGTTYSVATYDSAVETIANEQGNRYTPSPVAFSTSERLIGDAKKNQAL 63
QY 95 KPGARIF-----SFDGKVLRLH-PTWPKSVWHSNDPGRRLTE-SYCETWREAPS 144
DB 64 NPKNTYFADAKRLIGRAFDDSVQKDKISWPKYV----ESNGQPLEVEVLEDEKTKTSFQ 119
QY 145 ATGQASSLL 153
DB 120 ---EISSMV 125

RESULT 15
NRRL_CHICK
ID NRRL_CHICK STANDARD; PRT; 312 AA.
AC P55806;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 01-NOV-1997 (Rel. 35, last annotation update)
DE NAD(P)(+)-arginine ADP-ribosyltransferase 1 precursor (EC 2.4.2.31)
DE (Mono(ADP-ribosyl)transferase 1) (AT1).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=WHITE LEGHORN; TISSUE=Bone marrow;
RX MEDLINE=95050487; Pubmed=7961658;
RA Tsuchiya M., Hara N., Yamada K., Osago H., Shimoyama M.;
RT Cloning and expression of cDNA for arginine-specific ADP-
RT ribosyltransferase from chicken bone marrow cells.*;
RL J. Biol. Chem. 269:27451-27457(1994).
CC -1- CATALYTIC ACTIVITY: NAD(+) + L-arginine = nicotinamide + N2-
CC (ADP-D-ribosyl)-L-arginine.
CC -1- SUBCELLULAR LOCATION: THE MATURE ENZYME IS PROBABLY SECRETED
CC FROM GRANULOCYTES INTO THE EXTRACELLULAR SPACE.
CC -1- SIMILARITY: BELONGS TO THE ARG-SPECIFIC ADP-RIBOSYLTRANSFERASE
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D31864; BAA06664.1; -.
DR InterPro: IPR000768; ART.
DR Pfam: PF01129; ART; 1.
DR PRINTS: PR00970; RIBTRNSFRASE.
DR PROSITE: PS01291; ART; 1.
KW transferase; Glycosyltransferase; NAD; Signal; Zymogen.
FT SIGNAL 1 20
FT PROPEP 21 31
FT CHAIN 32 266
FT ACT_SITE 267 312
FT ACT_SITE 224 224
SQ SEQUENCE 312 AA; 35318 MW; B82360439BC904FC CRC64;

Query Match 7.3%; Score 70.5; DB 1; Length 312;
Best Local Similarity 24.8%; Pred. No. 13;
Matches 30; Conservative 17; Mismatches 45; Indels 29; Gaps 5;

QY 2 SHRDQPVLLHLVALNSPLSGMGRIGADFCQQAQAV-----GLAGTFRAFLSS 52
DB 128 SHQDYHSHYHKTLHFLTLQALFALRASQPCRYVYGVGRGIRPMQGRKSVRFQGTST 187
QY 53 RL-----QDLXSVRRADRAAVPIVNL-----KDELFPSEWAL-----FSGSGGPL 94
DB 188 SLRKEATVNGQDTLFFVKKTC-XGVPIKQSFSPSEDEVLIPPEFVEYINFSNDGVS 245
QY 95 K 95
DB 246 K 246

```

Search completed: August 12, 2002, 10:14:43  
 Job time: 130 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:11:43 ; Search time 16.76 Seconds

(without alignments)  
1049.185 Million cell updates/sec

Title: US-10-080-797-1

Perfect score: 967

Sequence: 1 HSHRDFQPVHLVALNSPLS.....SCHHATVLCIENSFWTAKS 183

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description               |
|------------|-------|-------------|--------|-------|---------------------------|
| 1          | 967   | 100.0       | 684    | 2     | A53019 collagen alpha 1(X |
| 2          | 840   | 86.9        | 1774   | 2     | B56101 collagen alpha 1(X |
| 3          | 837   | 86.6        | 1315   | 2     | A56101 collagen alpha 1(X |
| 4          | 529   | 54.7        | 1388   | 2     | A53317 collagen alpha 1(X |
| 5          | 373   | 38.6        | 650    | 2     | T22002 hypothetical prote |
| 6          | 79    | 8.2         | 477    | 2     | T05202 pectinesterase hom |
| 7          | 477   | 8.0         | 314    | 2     | F98231 succinoglycan bios |
| 8          | 77    | 8.0         | 314    | 2     | AG3054 succinoglycan bios |
| 9          | 76    | 7.9         | 204    | 2     | B45022 CRK-I - human      |
| 10         | 76    | 7.9         | 636    | 2     | JW0047 class I cytochrome |
| 11         | 75.5  | 7.8         | 904    | 2     | T04377 probable alpha-dex |
| 12         | 75    | 7.8         | 351    | 2     | B89781 conserved hypotet  |
| 13         | 74.5  | 7.7         | 850    | 2     | S56015 gastric mucin MUC5 |
| 14         | 74    | 7.7         | 1063   | 2     | T46284 hypothetical prote |
| 15         | 73.5  | 7.6         | 1653   | 2     | B91052 hypothetical prote |
| 16         | 73.5  | 7.6         | 1653   | 2     | G65028 hypothetical prote |
| 17         | 73.5  | 7.6         | 1653   | 2     | F85896 hypothetical prote |
| 18         | 73    | 7.5         | 302    | 2     | C71332 probable smf prote |
| 19         | 73    | 7.5         | 435    | 2     | JC5584 hyaluronoglycosam  |
| 20         | 73    | 7.5         | 462    | 2     | T17480 endo-xylanase homo |
| 21         | 73    | 7.5         | 513    | 1     | A45333 exopolyposphatase  |
| 22         | 73    | 7.5         | 513    | 2     | D91049 exopolyposphatase  |
| 23         | 73    | 7.5         | 513    | 2     | H85893 UDP-N-acetylglucos |
| 24         | 72.5  | 7.5         | 356    | 2     | AH0068 hypothetical prote |
| 25         | 72.5  | 7.5         | 841    | 2     | B71212 hypothetical prote |
| 26         | 72.5  | 7.5         | 884    | 2     | C70729 epidermal growth f |
| 27         | 72    | 7.4         | 239    | 2     | A46243 carbohydrate kins  |
| 28         | 72    | 7.4         | 297    | 2     | AE3128 CRK-II - human     |
| 29         | 72    | 7.4         | 304    | 2     | AA5022                    |

|    |      |     |      |   |                            |
|----|------|-----|------|---|----------------------------|
| 30 | 72   | 7.4 | 304  | 2 | I58394 c-Crk - mouse       |
| 31 | 72   | 7.4 | 319  | 2 | D98159 hypothetical prote  |
| 32 | 72   | 7.4 | 494  | 2 | A83204 arginate biosynthe  |
| 33 | 72   | 7.4 | 800  | 2 | AG0438 probable Rhs acces  |
| 34 | 72   | 7.4 | 1006 | 2 | T00050 hypothetical prote  |
| 35 | 71.5 | 7.4 | 554  | 2 | T06374 probable pectinest  |
| 36 | 71.5 | 7.4 | 554  | 2 | T06468 pectinesterase (EC  |
| 37 | 71.5 | 7.4 | 3670 | 2 | T36249 CDA peptide synthet |
| 38 | 71   | 7.3 | 579  | 2 | A70954 hypothetical prote  |
| 39 | 71   | 7.3 | 644  | 2 | A72519 probable 2-oxoacid  |
| 40 | 71   | 7.3 | 732  | 2 | AF0439 probable Rhs acces  |
| 41 | 71   | 7.3 | 787  | 2 | G81692 inner membrane pro  |
| 42 | 71   | 7.3 | 1154 | 2 | S43275 hypothetical prote  |
| 43 | 70.5 | 7.3 | 312  | 2 | A55461 MAD(P)-+-arginine   |
| 44 | 70.5 | 7.3 | 628  | 2 | D90677 probable propionyl  |
| 45 | 70.5 | 7.3 | 628  | 2 | G85527 probable propionyl  |

#### ALIGNMENTS

RESULT 1  
A53019 collagen alpha 1(XVIIII) chain - human (fragment)  
N:Contains: endostatin  
C:Species: Homo sapiens (man)  
C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 31-Mar-2000  
C/Accession: A53019  
R/OH, S.P.; Warman, M.L.; Seldin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olse  
Genomics 19, 494-499, 1994  
A>Title: Cloning of cDNA and genomic DNA encoding human type XVIIII collagen and local  
A/Reference number: A53019; MIM:94245237  
A/Accession: A53019  
A/Molecule type: mRNA  
A/Residues: 1-684 <OHA>  
A/Cross-references: GB:122548; NID:9348908; PIDN:AA51864.1; PID:9562794  
A/Note: The cited accession number, 122548, is not in Genbank release 103  
A/Note: In the authors' translation, 482-Gly is not shown, residues 483-490 are shift  
C/Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
laced and subsequently O-glycosylated.  
C/Comment: Different splice forms of collagen alpha 1(XVIIII) may be involved in periv  
C/Comment: Endostatin is released from collagen alpha 1(XVIIII) chain by the action of  
ay be useful in treating solid tumors.  
C/Genetics:  
A/Genes: GDB:COL18A1  
A/Cross-references: GDB:138752; OMIM:120328  
A/Map position: 21q22.3-21q22.3  
C/Superfamily: unassigned collagens  
C/Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog  
F:1-684/Product: collagen alpha 1(XVIIII) chain (fragment) #status predicted <MAT>  
F:1-59/Domain: collagenous (fragment) #status predicted <CO4>  
F:74-115/Domain: collagenous #status predicted <CO5>  
F:129-201/Domain: collagenous #status predicted <CO6>  
F:212-244/Domain: collagenous #status predicted <CO7>  
F:257-278/Domain: collagenous #status predicted <CO8>  
F:262-264/Region: cell attachment (R-G-D) motif  
F:286-340/Domain: collagenous #status predicted <CO9>  
F:354-371/Domain: collagenous #status predicted <CO10>  
F:502-684/Product: endostatin #status predicted <EST>  
F:509-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 100.0%; Score 967; DB 2; Length 684;  
Best Local Similarity 100.0%; Pred. No. 3.2e-88;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 HSHRDFQPVHLVALNSPLSGMGIRGADFOCFQQAARAVGLAGTFRAFSSRLQDLYSI 60  
DB 502 HSHRDFQPVHLVALNSPLSGMGIRGADFOCFQQAARAVGLAGTFRAFSSRLQDLYSI 561  
QY 61 VRRADRAAVPIVNLKDELFPSSWEALFSSGSEGLPKGARIFSEDKVLRHPTWPKSW 120  
DB 562 VRRADRAAVPIVNLKDELFPSSWEALFSSGSEGLPKGARIFSEDKVLRHPTWPKSW 621

QY 121 HSDPGRRLTESYCTWTEAPSGASLLGRLGQSAASHHAYIVLCIENSFMT 180  
|||||  
DB 622 HSDPGRRLTESYCTWTEAPSGASLLGRLGQSAASHHAYIVLCIENSFMT 681  
QY 181 ASK 183  
|||  
DB 682 ASK 684  
RESULT 2  
856101  
collagen alpha 1(XVIII) chain precursor, long splice form - mouse  
N:Contains: collagen alpha 1(XVIII) chain precursor, medium splice form; endostatin  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Oct-1995 #sequence-revision 08-May-1998 #text-change 31-Mar-2000  
A:Accession: B56101; C56101; S72450; S65595; P06675; A54072; A58816  
R:Rehm, M.; Pihlajaniemi, T.  
J. Biol. Chem. 270, 4705-4711, 1995  
A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue homologous to rat and Drosophila frizzled proteins.  
A:Reference number: A56101; MUID:95181468  
A:Accession: B56101  
A:Molecule type: mRNA  
A:Residues: 1-562 <REH1>  
A:Cross-references: GB:U11637; NID:9618429; PIDN:AA52179.1; PID:9618430  
A:Experimental source: splice form clone PE17.24  
A:Accession: C56101  
A:Molecule type: mRNA  
A:Residues: 1-239, 487-562 <REH2>  
A:Cross-references: GB:U11637; NID:9618429  
A:Experimental source: splice form clones PE8.1, PE19, PE15.2  
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
submitted to the EMBL Data Library, August 1993  
A:Reference number: S72450  
A:Accession: S72450  
A:Molecule type: mRNA  
A:Residues: 487-1146, 'L', 1148-1193, 'F', 1195-1210, 'R', 1212-1512, 'L', 1514-1522, 'F', 1524-16  
A:Cross-references: EMBL:L22945; NID:9348968; PIDN:AA19787.1; PID:9511298  
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa  
A:Reference number: A58370; MUID:94240111  
A:Accession: S65595  
A:Molecule type: mRNA  
A:Residues: 487-1512, 'L', 1514-1522, 'F', 1524-1683, 'V', 1685-1774 <OH2>  
A:Cross-references: EMBL:L22545  
R:Abu, N.; Muragaki, Y.; Yoshioke, H.; Inoue, H.; Minomiy, Y.  
Biochem. Biophys. Res. Commun. 196, 576-582, 1993  
A:Title: Identification of a novel collagen chain represented by extensive interruptions  
A:Reference number: P06675; MUID:94059075  
A:Accession: P06675  
A:Molecule type: mRNA  
A:Residues: 635-1774 <ABE>  
R:Rehm, M.; Hintikka, E.; Pihlajaniemi, T.  
J. Biol. Chem. 269, 13929-13935, 1994  
A:Title: Primary structure of the alpha1 chain of mouse type XVIII collagen, partial str  
collagen chain.  
A:Reference number: A54072; MUID:94245707  
A:Accession: A54072  
A:Molecule type: DNA  
A:Residues: 1293-1403, 'R', 1405-1774 <REH3>  
A:Cross-references: GB:U03714; NID:9487733; PIDN:AA20657.1; PID:9487734  
R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukui, N.; Vasios, G.; Lane, W.S.; Flynn, E.; B  
Cell 88, 277-285, 1997  
A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.  
A:Reference number: A58816; MUID:97160848  
A:Accession: A58816  
A:Molecule type: protein  
A:Residues: 1591-1610 <ORE>  
A:Experimental source: hemangioendothelium cells  
A:Note: Inhibits endothelial cell proliferation  
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit

lated and subsequently O-glycosylated.  
C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in p  
C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of  
ay be useful in treating solid tumors.

C:Genetics:  
A:Gene: MGI:coll18a1  
A:Cross-references: MGI:71175  
A:Map position: 10:41.0  
A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 15  
A:Note: The list of introns is incomplete  
C:Superfamily: unassigned collagens  
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteog  
F:1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status p  
F:1-239,487-1774/Product: collagen alpha 1(XVIII) chain precursor, medium splice form  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:786-486/Region: frizzled similarity  
F:786-812/Domain: collagenous #status predicted <CO01>  
F:823-896/Domain: collagenous #status predicted <CO02>  
F:921-1042/Domain: collagenous #status predicted <CO03>  
F:1066-1148/Domain: collagenous #status predicted <CO04>  
F:1163-1204/Domain: collagenous #status predicted <CO05>  
F:1218-1290/Domain: collagenous #status predicted <CO06>  
F:1301-1333/Domain: collagenous #status predicted <CO07>  
F:1346-1369/Domain: collagenous #status predicted <CO08>  
F:1351-1353/Region: cell attachment (R-G-D) motif  
F:1377-1428/Domain: collagenous #status predicted <CO09>  
F:1442-1459/Domain: collagenous #status predicted <CO10>  
F:1591-1774/Product: endostatin #status predicted <EST>  
F:1598-1774/Region: multiplexin collagen carboxyl-terminal similarity  
F:354,361,947/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:699,704,1716/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted  
Query Match 86.9%; Score 840; DB 2; Length 1774;  
Best Local Similarity 85.6%; Pred. No. 4,6e-75;  
Matches 155; Conservative 15; Mismatches 11; Indels 0; Gaps 0;  
QY 1 HSHRDFQVPLHVALNPLSGMGRIGRADFCQFQARAVGLAGTFRAFLSSRLDLYSI 60  
|||||  
DB 1591 HHODEQVPLHVALNPLSGMGRIGRADFCQFQARAVGLSGFRAPLSSRLDLYSI 1650  
QY 61 VRRADRAAPVYNLDELFPSEALFSGSEGLPKGAISEFDGKDVLRHPTWQKSW 120  
|||||  
DB 1651 VRRADRGVPIYNLDELVPSPWDSLFSSGQGLDPGARIFFEDGVDVLRHPAMPQKSW 1710  
QY 121 HSDPGRRLTESYCTWTEAPSGASLLGRLGQSAASHHAYIVLCIENSFMT 180  
|||||  
DB 1711 HSDPGRRLTESYCTWTEAPSGASLLGRLGQSAASHHAYIVLCIENSFMT 1770  
QY 181 A 181  
DB 1771 S 1771  
RESULT 3  
collagen alpha 1(XVIII) chain precursor, short splice form - mouse  
N:Contains: endostatin  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Oct-1995 #sequence-revision 08-May-1998 #text-change 31-Mar-2000  
A:Accession: A56101; A58371; S72450; S65595  
R:Rehm, M.; Pihlajaniemi, T.  
J. Biol. Chem. 270, 4705-4711, 1995  
A:Title: Identification of three N-terminal ends of type XVIII collagen chains and ti  
t homologous to rat and Drosophila frizzled proteins.  
A:Reference number: A56101; MUID:95181468  
A:Accession: A56101  
A:Molecule type: mRNA  
A:Residues: 1-103 <REH1>  
A:Cross-references: GB:U11636; NID:9618427; PIDN:AA52178.1; PID:9618428  
R:Rehm, M.; Pihlajaniemi, T.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994

A:Title: Alpha1(XVII), a collagen chain with frequent interruptions in the collagenous  
A:Reference number: A58371; MUID:94240112  
A:Accession: A58371  
A:Molecule type: mRNA  
A:Residues: 1-928 <REH>  
A:Cross-references: GB:L16898; NID:9404754; PIDN:AAA37434.1; PID:9553994  
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
Submitted to the EMBL Data Library, August 1993  
A:Reference number: S72450  
A:Accession: S72450  
A:Molecule type: mRNA  
A:Residues: 28-687, 'L', 689-734, 'F', 736-751, 'R', 753-1315 <OHM>  
A:Cross-references: EMBL:L22545; NID:9348968; PIDN:AA19787.1; PID:9511298  
R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa  
A:Reference number: A58370; MUID:94240111  
A:Accession: S65595  
A:Molecule type: mRNA  
A:Residues: 28-1315 <OHS>  
A:Cross-references: EMBL:L22545  
A:Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
lated and subsequently O-glycosylated.  
C:Comment: The different splice forms of collagen alpha 1(XVII) may be involved in per  
C:Comment: Endostatin is released from collagen alpha 1(XVII) chain by the action of un  
ay be useful in treating solid tumors.  
C:Genetics:  
A:Gene: MGI:COL18A1  
A:Cross-references: MGI:71175  
A:Map position: 10:41.0  
C:Superfamily: unassigned collagens  
C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:24-235/Region: thrombospondin amino-terminal similarity  
F:26-1315/Product: collagen alpha 1(XVII) chain, short splice form #status predicted <  
F:327-353/Domain: collagenous #status predicted <CO1>  
F:364-437/Domain: collagenous #status predicted <CO2>  
F:462-583/Domain: collagenous #status predicted <CO3>  
F:607-689/Domain: collagenous #status predicted <CO4>  
F:704-745/Domain: collagenous #status predicted <CO5>  
F:759-831/Domain: collagenous #status predicted <CO6>  
F:842-874/Domain: collagenous #status predicted <CO7>  
F:887-910/Domain: collagenous #status predicted <CO8>  
F:892-894/Region: cell attachment (R-G-D) motif  
F:918-966/Domain: collagenous #status predicted <CO9>  
F:983-1000/Domain: collagenous #status predicted <CO10>  
F:1132-1315/Product: endostatin #status predicted <EST>  
F:1133-1315/Region: multiplexin collagen carboxyl-terminal similarity  
F:126,488/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F:172-228/Diulfide bonds: #status predicted  
F:240,245,1257/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:451,454,594/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 86.6%; Score 837; DB 2; Length 1315;  
Best Local Similarity 85.1%; Pred. No. 6,4e-75;  
Matches 154; Conservative 16; Mismatches 11; Indels 0; Gaps 0;

QY 1 HSRDQPVHLALVNLSPISGGRGIRGADFOCFQARAVGLAGTFRALSSRLQDLYST 60  
DB 1132 HTIQDQVPLHLVNLTPISGGRGIRGADFOCFQARAVGLAGTFRALSSRLQDLYST 1191

QY 61 VRRADRAVPVIVLKDDELFPSSNEALFSGSEGLPKGARIFSGDKVLRHPTWPKSVW 120  
DB 1192 VRRADRAVPVIVLKDDELFPSSNEALFSGSEGLPKGARIFSGDKVLRHPTWPKSVW 1251

QY 121 HGSDFNGRRLTESYCTWTREAPATGQASLLGRLLAGSAAASHAHAYIVLCIENSFMT 180  
DB 1252 HGSDFNGRRLTESYCTWTREAPATGQASLLGRLLAGSAAASHAHAYIVLCIENSFMT 1311

QY 181 A 181  
DB 1312 S 1312

RESULT 4  
A53317  
collagen alpha 1(XV) chain precursor - human  
N:Alternate names: procollagen alpha 1(XV) chain  
C:Species: Homo sapiens (hmn)  
C>Date: 07-Jul-1995 #sequence  
C:Accession: A53317; S28778  
R:Kivirikko, S.; Heinaemeaki, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, J. Biol. Chem. 269, 4773-4779, 1994  
A:Title: Primary structure of the alpha1 chain of human type XV collagen and exon-int  
A:Reference number: A53317; MUID:94148920  
A:Accession: A53317  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1388 <KITV>  
A:Cross-references: GB:L25280  
A:Note: nucleotide sequence and conceptual translation not complete  
R:Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A.  
J. Biol. Chem. 269, 4042-4046, 1994  
A:Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-trip  
A:Reference number: A53146; MUID:94140817  
A:Accession: A53146  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-9, 'S', 11-48, 'V', 50-94, 'A', 96-149, 'A', 151-203, 'V', 205-408, 'A', 410-569 <M  
A:Cross-references: GB:D21230; NID:9415605; PIDN:BA04762.1; PID:d1005294; PID:946070  
R:Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.  
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992  
A:Title: Identification of a previously unknown human collagen chain, alpha1(XV), cha  
A:Reference number: S28778; MUID:93066196  
A:Accession: S28778  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 544-640, 'P', 642-811, 'P', 813-1252 <WYE>  
A:Genetics:  
A:Gene: GDB:COL15A1  
A:Cross-references: GDB:132578; OMIM:120325  
A:Map position: 9q21-9q22  
C:Superfamily: unassigned collagens  
C:Keywords: signal sequence #status predicted <SIG>  
F:1-22/Domain: signal sequence #status predicted <MAT>  
F:23-1388/Product: collagen alpha 1(XV) chain #status predicted  
F:1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 54.7%; Score 529; DB 2; Length 1388;  
Best Local Similarity 55.9%; Pred. No. 3,4e-44;  
Matches 99; Conservative 30; Mismatches 44; Indels 4; Gaps 1;

QY 7 QPVHLVNLSPISGGRGIRGADFOCFQARAVGLAGTFRALSSRLQDLYSTVRRADR 66  
DB 1215 KPALHIALNPFSGDIR---ADFQCFQARAAGLLSTYRAFLSSHQDLYSTVRAER 1270

QY 67 AAVPVLVNLKDDELFPSSNEALFSGSEGLPKGARIFSGDKVLRHPTWPKSVWHSQDPN 126  
DB 1271 YSLPVLVNLKQGVLENMNDISFSGHGGGFNMHPIRYSFGDRIMDPMPQKVIHNGSPH 1330

QY 127 GRRLTESYCTWTREAPATGQASLLGRLLAGSAAASHAHAYIVLCIENSFMTASK 183  
DB 1331 GVRLLVDMVCEAMRTRADPAVGLASPLSTGKITDOKAVSCANRLIVLCIENSFMTDAR 1387

RESULT 5  
T22002  
hypothetical protein F39H11.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence  
C:Accession: T22002  
R:White, S.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19500  
A:Accession: T22002





R:Matsumura, M.; Tanaka, S.; Nagata, S.; Kojima, A.; Kurata, T.; Shibuya, M.  
Mol. Cell. Biol. 12, 3482-3489, 1992  
A:Title: Two species of human CRK cDNA encode proteins with distinct biological activities  
A:Reference number: A45022; MID:92334347  
A:Accession: B45022  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-204 <MAT>  
A:Experimental source: embryonic lung cells  
A:Note: sequence extracted from NCBI backbone (NCBIN:108771, NCBI:108772)  
C:Superfamily: crk transforming protein, SH2 homology; SH3 homology  
F:13-118/Domain: SH2 homology <SH2>  
F:139-187/Domain: SH3 homology <SH3>

Query Match 7.9%; Score 76; DB 2; Length 204;  
Best Local Similarity 29.0%; Pred. No. 4.6;  
Matches 20; Conservative 12; Mismatches 21; Indels 16; Gaps 3;

QY 100 IFSPDK-----DVLHPTWPOKSWHSGSDPNRR--LTESYCEITRTEAPSGATG 147  
DB 140 LFPNGNDEHDLFPKGDILIRDKPEQWMADESGKGMIPVYKRR----PASA 195  
QY 148 QASLLIGR 156  
DB 196 SVSALIGR 204

RESULT 10  
JM0047  
Class I cytokinase receptor precursor - human  
N:Alternate names: MSX-1  
C:Species: Homo sapiens (man)  
C:Date: 17-Jun-1998 #sequence\_revision 10-Jul-1998 #text\_change 21-Jul-2000  
C:Accession: JM0047  
R:Spieker, C.A.; Grant, F.J.; Baumgartner, J.W.; Presnell, S.R.; Schrader, S.K.; Yamaguchi, R.; Blochem, B.; Res. Commun. 246, 82-90, 1998  
A:Title: Cloning and characterization of a novel class I cytokine receptor.  
A:Reference number: JM0047; MID:98262921  
A:Accession: JM0047  
A:Molecule type: mRNA  
A:Residues: 1-636 <SPR>  
A:Cross-References: GB:AF053004; MID:g3153240; PIDN:AAC39755.1; PID:g3153241  
A:Experimental source: brain  
C:Genetics:

A:Map position: 19p13.11  
C:Keywords: glycoprotein  
F:1-33/Domain: signal sequence #status predicted <SIG>  
F:515-540/Domain: transmembrane #status predicted <TM>  
F:554-561/Domain: cytoplasmic #status predicted <CTP>  
F:51,76,302,311,374,382,467/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.9%; Score 76; DB 2; Length 636;  
Best Local Similarity 24.2%; Pred. No. 18;  
Matches 39; Conservative 12; Mismatches 46; Indels 64; Gaps 7;

QY 23 MRIRADP-----QCFQARAVGLAGTFRAFLSSRLDL----- 57  
DB 1 MRGRGAPFWLWPLKIALPLMLVLFQRTTPGSGAPLOCYGVGLDNCSEPLGLD 60  
QY 58 -----YSIVRRADRAAVPI-----VNLKDELLF-----PSWEALFS 88  
DB 61 GAPSEHLLOSQRKSKTKYVAVAGRSWVAITREQLTHSDKLLWGTAKAGPLMPVYFV 120  
QY 89 GSEGPLKPGA-RI--FSPDKDVLK-----HPTWPOKSV 119  
DB 121 NLETOMKPNAPRLGPDVDFSEDDPLEATVHAMDPPTWPSHKV 161

RESULT 11  
T04377  
probable alpha-dextrin endo-1,6-alpha-glucosidase (EC 3.2.1.41) - barley

N:Alternate names: pullulanase  
C:Species: Hordeum vulgare (barley)  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 08-Oct-1999  
C:Accession: T04377  
R:Lock, F.; Kristensen, M.; Planchot, V.; Leah, R.; Svendsen, I.; Svenson, B.  
submitted to the EMBL Data Library, December 1997  
A:Description: Isolation and characterization of starch debranching enzyme, limit dex  
A:Reference number: Z15320  
A:Accession: T04377  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-904 <LOK>  
A:Cross-References: EMBL:AF022725; MID:g2502057; PIDN:AD04189.1; PID:g2677837  
A:Experimental source: cv. Igri  
C:Genetics:  
A:Gene: HVD99  
A:Introns: 21/3; 70/1; 87/3; 124/1; 169/3; 235/2; 285/2; 305/3; 346/1; 376/2; 407/3;  
C:Keywords: glycosidase; hydrolase

Query Match 7.8%; Score 75.5; DB 2; Length 904;  
Best Local Similarity 22.3%; Pred. No. 31;  
Matches 45; Conservative 19; Mismatches 67; Indels 69; Gaps 8;

QY 17 SPLSGMRGIRGADFOCFQARAVGLAGT-----PRAFLSSRLDYSIVR---- 62  
DB 62 SPNSG-----GIQGYDSKELPESAGLEPTVYQKFPFISYAFVPSVYASLVKQOLV 118  
QY 63 ----RADRAAVPIVNL-----KDELLFPSWE-----ALFSG 89  
DB 119 VASFGADGKHVDYGLQLPGVLDPMFAVTGPGVAFSEDSVSLHLMPTAGVSVCFPDG 178  
QY 90 SEGP-----LKPARRFSPDK-----DVLHPTWPOKSWHSGSDPNRRLT 131  
DB 179 PAGALPTVQLKESNGVSWGVPREWMENRYLYEVYV--HPTKAQVLKCLAGDPYASLS 237  
QY 132 ESYCEITRTEAPSGAASS 151  
DB 238 ANGARTLVLDINNETLKPAS 257

RESULT 12  
B89781  
conserved hypothetical protein SA0184 [Imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: B89781  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
ma, A.; Mizutani, U.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MID:21311952; PMID:11418146  
A:Accession: B89781  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1351 <KUR>  
A:Cross-References: GB:BA000018; PID:g13700106; PIDN:BA041405.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0184

Query Match 7.8%; Score 75; DB 2; Length 351;  
Best Local Similarity 23.3%; Pred. No. 11;  
Matches 42; Conservative 17; Mismatches 67; Indels 54; Gaps 8;

QY 23 MRGIRGADFOCFQARAVGLAGTFRAFLSSRLDL-----YSIVRRADRA-AVPIV 72  
DB 97 IEAIMAOGKCCUNASTITS-----RELITSLDQDLNFTLISFCHNYVPRDGLSLDVLV 151  
QY 73 NLKDELLFPSWEALFSGSEGPLKPGARIFSPDKDVLHPTWPOKSWHSGSDPNRRLT 132

Db 152 NKKNELLY-----QFNKKAIIYGTIVSSGLRGLP-----HKGLP----- 186

QY 133 SYCEWRTPEA-----SATGQASSLLGCGLLGQSA-----CHHATVLCIENSFMT 180

Db 187 --IEATRSHVVAAKLLQETGCVSEVLYGSLIEMRQAKQILDFCKRHHTTLCIEEFPDI 244

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:11:08 ; Search time 13.23 seconds  
(without alignments)  
337.860 Million cell updates/sec

Title: US-10-080-797-1  
Perfect score: 967  
Sequence: 1 HSHRDFQVFLHLVNLSPLS.....SCHHAYIVLCIENSFWTASK 183

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/1aa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/1aa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID               | Description       |
|------------|-------|-------------|--------|------------------|-------------------|
| 1          | 967   | 100.0       | 183    | US-09-206-059-2  | Sequence 2, Appl  |
| 2          | 962   | 99.5        | 182    | US-09-561-500-14 | Sequence 14, Appl |
| 3          | 962   | 99.5        | 182    | US-09-561-108-14 | Sequence 14, Appl |
| 4          | 962   | 99.5        | 182    | US-09-315-689-3  | Sequence 3, Appl  |
| 5          | 937   | 96.9        | 178    | US-09-315-689-5  | Sequence 5, Appl  |
| 6          | 840   | 86.9        | 191    | US-09-561-500-13 | Sequence 13, Appl |
| 7          | 840   | 86.9        | 191    | US-09-561-108-13 | Sequence 13, Appl |
| 8          | 837   | 86.6        | 195    | US-08-159-784-2  | Sequence 2, Appl  |
| 9          | 796   | 82.3        | 185    | US-08-985-526-36 | Sequence 36, Appl |
| 10         | 516   | 53.4        | 191    | US-08-159-784-3  | Sequence 3, Appl  |
| 11         | 160   | 16.5        | 35     | US-09-046-985-2  | Sequence 2, Appl  |
| 12         | 160   | 16.5        | 35     | US-09-474-743-7  | Sequence 7, Appl  |
| 13         | 101   | 10.4        | 22     | US-09-046-985-7  | Sequence 7, Appl  |
| 14         | 101   | 10.4        | 22     | US-09-474-743-7  | Sequence 7, Appl  |
| 15         | 96    | 9.9         | 20     | US-08-740-168A-1 | Sequence 1, Appl  |
| 16         | 96    | 9.9         | 20     | US-09-349-429-1  | Sequence 1, Appl  |
| 17         | 96    | 9.9         | 20     | US-09-315-689-1  | Sequence 1, Appl  |
| 18         | 94    | 9.7         | 16     | US-09-385-442-32 | Sequence 32, Appl |
| 19         | 76    | 7.9         | 578    | US-08-653-740-3  | Sequence 3, Appl  |
| 20         | 76    | 7.9         | 578    | US-09-073-594-3  | Sequence 3, Appl  |
| 21         | 76    | 7.9         | 578    | US-09-275-925-3  | Sequence 3, Appl  |
| 22         | 76    | 7.9         | 636    | US-08-653-740-5  | Sequence 5, Appl  |
| 23         | 76    | 7.9         | 636    | US-09-073-594-5  | Sequence 5, Appl  |
| 24         | 76    | 7.7         | 256    | US-09-275-925-5  | Sequence 5, Appl  |
| 25         | 74    | 7.7         | 256    | US-07-906-349A-8 | Sequence 8, Appl  |
| 26         | 74    | 7.7         | 256    | US-08-167-035-4  | Sequence 4, Appl  |
| 27         | 74    | 7.7         | 256    | US-08-208-887A-4 | Sequence 4, Appl  |

|    |      |     |      |                   |                    |
|----|------|-----|------|-------------------|--------------------|
| 28 | 74   | 7.7 | 256  | US-08-539-005-4   | Sequence 4, Appl   |
| 29 | 73   | 7.5 | 311  | US-08-987-743-2   | Sequence 2, Appl   |
| 30 | 73   | 7.5 | 435  | US-08-733-360A-1  | Sequence 1, Appl   |
| 31 | 73   | 7.5 | 435  | US-08-733-360A-3  | Sequence 3, Appl   |
| 32 | 73   | 7.5 | 435  | US-08-987-743-6   | Sequence 6, Appl   |
| 33 | 73   | 7.5 | 435  | US-08-987-743-15  | Sequence 15, Appl  |
| 34 | 73   | 7.5 | 435  | US-08-916-935-1   | Sequence 1, Appl   |
| 35 | 73   | 7.5 | 435  | US-08-916-935-3   | Sequence 3, Appl   |
| 36 | 72   | 7.4 | 304  | US-08-630-915A-28 | Sequence 28, Appl  |
| 37 | 70   | 7.2 | 210  | US-09-188-930-148 | Sequence 148, Appl |
| 38 | 70   | 7.2 | 6095 | US-09-144-085-2   | Sequence 2, Appl   |
| 39 | 69   | 7.1 | 190  | US-09-046-985-15  | Sequence 15, Appl  |
| 40 | 69   | 7.1 | 190  | US-09-474-743-15  | Sequence 15, Appl  |
| 41 | 69   | 7.1 | 313  | US-09-347-803-6   | Sequence 6, Appl   |
| 42 | 69   | 7.1 | 327  | US-09-168A-6      | Sequence 6, Appl   |
| 43 | 68.5 | 7.1 | 1289 | US-07-876-280-4   | Sequence 4, Appl   |
| 44 | 68.5 | 7.1 | 1289 | US-07-675-772-4   | Sequence 4, Appl   |
| 45 | 68.5 | 7.1 | 1289 | US-08-063-170-4   | Sequence 4, Appl   |

ALIGNMENTS

```
RESULT 1
US-09-206-059-2
; Sequence 2, Application US/09206059
; Patent No. 6201104
; GENERAL INFORMATION:
; APPLICANT: Macdonald, Nicholas
; APPLICANT: Sim, Kim Lee
; TITLE OF INVENTION: Angiogenesis-Inhibiting Protein Binding Peptides and
; FILE REFERENCE: 05213-0370
; CURRENT APPLICATION NUMBER: US/09/206,059
; CURRENT FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-206-059-2

Query Match      100.0%; Score 967; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 2e+110;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 HSHRDFQVFLHLVNLSPLSGMRGIRGADFCQFOARAVGLAGTFRFLSSRLQDLXSI 60
      |||
DB      1 HSHRDFQVFLHLVNLSPLSGMRGIRGADFCQFOARAVGLAGTFRFLSSRLQDLXSI 60

QY      61 VRRADRAVPIVNLKDELLFPSEWELFSGSEGPLKPGARIFSPDKDYLRRHPWPQKSVW 120
      |||
DB      61 VRRADRAVPIVNLKDELLFPSEWELFSGSEGPLKPGARIFSPDKDYLRRHPWPQKSVW 120

QY      121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFWT 180
      |||
DB      121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFWT 180

QY      181 ASK 183
      |||
DB      181 ASK 183

RESULT 2
US-09-561-500-14
; Sequence 14, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Philip E. Thorpe
; APPLICANT: Rolf A. Brecken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
```

```
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-500-14
```

```
Query Match          99.5%; Score 962; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 8.2e-110;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 HSHRDPQVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLQDLYSI 60
DB 1 HSHRDPQVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLQDLYSI 60
QY 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
QY 121 HGSDDPNGRRLTESYCETWRTAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
DB 121 HGSDDPNGRRLTESYCETWRTAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
QY 181 AS 182
DB 181 AS 182
```

```
RESULT 3
US-09-561-108-14
; Sequence 14, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Phillip E. Thorpe
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 14
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
US-09-561-108-14
```

```
Query Match          99.5%; Score 962; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 8.2e-110;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 HSHRDPQVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLQDLYSI 60
DB 1 HSHRDPQVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLQDLYSI 60
QY 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
```

```
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
QY 121 HGSDDPNGRRLTESYCETWRTAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
DB 121 HGSDDPNGRRLTESYCETWRTAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
QY 181 AS 182
DB 181 AS 182
```

```
RESULT 4
US-09-315-689-3
; Sequence 3, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-3
```

```
Query Match          99.5%; Score 962; DB 4; Length 182;
Best Local Similarity 100.0%; Pred. No. 8.2e-110;
Matches 182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 HSHRDPQVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLQDLYSI 60
DB 1 HSHRDPQVHLVALNSPLSGMKGIRGADFCFOQARAVGLAGTFRATLSSRLQDLYSI 60
QY 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
DB 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGPLKPGARIFSPDGKDVLRHPTWPKSVW 120
QY 121 HGSDDPNGRRLTESYCETWRTAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
DB 121 HGSDDPNGRRLTESYCETWRTAPSATGQASSLLGRLGQSAASHHAYIVLCIENSFMT 180
QY 181 AS 182
DB 181 AS 182
```

```
RESULT 5
US-09-315-689-5
; Sequence 5, Application US/09315689
; Patent No. 6346510
; GENERAL INFORMATION:
; APPLICANT: Folkman, Judah
; APPLICANT: O'Reilly, Michael
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions
; FILE REFERENCE: 05213-0229
; CURRENT APPLICATION NUMBER: US/09/315,689
; CURRENT FILING DATE: 1999-05-20
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-315-689-5
```

```
Query Match          96.9%; Score 937; DB 4; Length 178;
```



US-08-159-784-2

|                       |                  |                    |           |             |
|-----------------------|------------------|--------------------|-----------|-------------|
| Query Match           | 86.68;           | Score 837;         | DB 1;     | Length 195; |
| Best Local Similarity | 85.18;           | Pred. NO. 1.8e-94; |           |             |
| Matches 154;          | Conservative 16; | Mismatches 11;     | Indels 0; | Gaps 0;     |

```

QY      1 HSHDFQVVLVNLVNLNSPLSGMGINGADPOCCQOARAVGLACTFAAFLSRLDDLYSI 60
Db      12 HTHDFQVVLVNLVNLNPPLSGMGINGADPOCCQOARAVGLSGTFAAFLSRLDDLYSI 71

QY      61 VRRADRAAVPIVNLKDELLFPSEWALEFSGSEGPLKPGARIFSPDGKDVLRHPTWPQKSW 120
Db      72 VRRADRGSPVIVNLKDEVLSPTSMDSLFPSGQGVQGVQPGARIFSPDGRVLRHNPMPQKSW 131

QY      121 HGSDPNGRRRLTESCEFWRTFAAPATGQASSLLGGRLGLGQSAAASHIAYTVLCIENSFMT 180
Db      132 HGSDPSPGRRRLMESYCEFWRTETTGATGQASSLLGRLLEQKAASHNSIYIVLCIENSFMT 191

QY      181 A 181
Db      192 S 192

```

```

: RESULT 9
: US-08-985-526-36
: Sequence 36, Application US/08985526
: Patent No. 6080728
: GENERAL INFORMATION:
: APPLICANT: MIXSON, James A
: TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA
: TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENETIC
: TITLE OF INVENTION: THERAPY
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Connolly, Bove, Lodge, & Hutz
: STREET: 1220 Market Street, P.O. Box 2207
: CITY: Wilmington
: STATE: Delaware
: COUNTRY: U.S.A.
: ZIP: 19899
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/985,526
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/608,845
: FILING DATE: 16-JUL-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: MCNORTON JR., Robert G
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (302) 658-9141
: TELEFAX: (302) 658-5613
: INFORMATION FOR SEQ ID NO: 36:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 185 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: US-08-985-526-36

```

```
Query Match      82.3%; Score 796; DB 3; Length 165;  
Best Local Similarity 82.4%; Pred. No. 1.7e-89;  
Matches 150; Conservative 16; Mismatches 14; Indels 2; Gaps 2.  
  
QY 1 HSHRFPQVLYLHVALNSPLSGCMRCIRGADPFCFOARAVAGLAGTFRRLSSRLQDYSI 60  
|:::|||||||:::|||||||:::|||||||:::|||||||:::|||||||
```

|    |     |  |     |
|----|-----|--|-----|
| Db | 2   | HTHDDFQVHLVLAINTPLSGMGNGICGADPQCENMAR-VGLSOTFAFLSLRLQDLYSI     | 60  |
| QY | 61  | VRRADRAAIVY-NLKDELLFPSEWEALFSGSEGPLKRGARIFSPDGKADVYRHPWPQKSV   | 119 |
| Db | 61  | VRRADRGVPIVQMLRDEVLSPSWDSLFGSGOGOLQGARIFSPDGDVYRHPAMPORSV      | 120 |
| QY | 120 | WHGSDPNCRRLTESYCEFTWRTFAPASATGOASSLLGRLGLQSSAASCHNAVIVLCIENSFM | 179 |
| Db | 121 | WHGSDPSCRRIMESYCEFTWRTTETTGATGOASSLLGRLGLQRAASCHDSIVLICIENSFM  | 180 |
| QY | 180 | TA 181   |     |
| Db | 181 | TS 182   |     |

```

RESULT 10
US-08-159-784-3
: Sequence 3, Application US/08159784
: Patent No. 5643783
:
GENERAL INFORMATION:
: APPLICANT: Bjorn R. Olsen
: TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: U.S.A.
: Zip: 02110-2804
:
COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: COMPUTER: IBM PS/2 Model 502 or 55SX
: OPERATING SYSTEM: MS-DOS (Version 5.0)
: SOFTWARE: Wordperfect (Version 5.1)
:
CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/159,784
: FILING DATE: December 1, 1993
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
:
ATTORNEY/AGENT INFORMATION:
: NAME: John F. Freeman
: REGISTRATION NUMBER: 29,066
: REFERENCE/DOCKET NUMBER: 00246/170001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 542-5070
: TELEFAX: (617) 542-8906
: TELEX: 200154
:
INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 191
: TYPE: amino acid
: STRANDEDNESS: N/A
: TOPOLOGY: N/A
:
US-08-159-784-3

```

[illegible]

Db 134 GVRLVNVCAMRTADTAVTGLASPLSTGKILDKAKVSCANRLIVLCIENSFMTDAR 190

RESULT 11

US-09-046-985-2

Sequence 2, Application US/09046985

Patent No. 6121236

GENERAL INFORMATION:

APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/046,985

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 861-9540

TELEFAX: (781) 861-6240

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 35 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-046-985-2

Query Match 16.5%; Score 160; DB 3; Length 35;

Best Local Similarity 100.0%; Pred. No. 9.9e-13;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 QARAVGLAGTFRALFLSRLODLYSIVRRADRAAV 69

Db 1 QARAVGLAGTFRALFLSRLODLYSIVRRADRAAV 34

RESULT 12

US-09-474-743-2

Sequence 2, Application US/09474743

Patent No. 6255716

GENERAL INFORMATION:

APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/046,985

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 861-9540

TELEFAX: (781) 861-6240

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/474,743

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION NUMBER:

APPLICATION NUMBER: 09/046,985

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 861-9540

TELEFAX: (781) 861-6240

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 35 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-474-743-2

Query Match 16.5%; Score 160; DB 4; Length 35;

Best Local Similarity 100.0%; Pred. No. 9.9e-13;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 36 QARAVGLAGTFRALFLSRLODLYSIVRRADRAAV 69

Db 1 QARAVGLAGTFRALFLSRLODLYSIVRRADRAAV 34

RESULT 13

US-09-046-985-7

Sequence 7, Application US/09046985

Patent No. 6121236

GENERAL INFORMATION:

APPLICANT: Ben-Sasson, Shmuel A.

TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: USA

ZIP: 02173

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/046,985

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Brook, David E.

REGISTRATION NUMBER: 22,592

REFERENCE/DOCKET NUMBER: CMCC-614

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781) 861-9540

TELEFAX: (781) 861-6240

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 22 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= modified aa  
OTHER INFORMATION: /note= "N-Acetyl-Threonine"  
US-09-046-985-7

Query Match 10.4%; Score 101; DB 3; Length 22;  
Best Local Similarity 100.0%; Pred. No. 8.4e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 45 TFRAPLSSRLQDLXSYIVRAD 65  
1 TFRAPLSSRLQDLXSYIVRAD 21

RESULT 14  
US-09-474-743-7  
Sequence 7, Application US/09474743  
Patent No. 6235716  
GENERAL INFORMATION:  
APPLICANT: Ben-Sasson, Shmuel A.  
TITLE OF INVENTION: NOVEL MULTIVALENT LIGANDS WHICH MODULATE  
TITLE OF INVENTION: ANGIOGENESIS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Millida Drive  
CITY: Lexington  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/474,743  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 09/046,985  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Brook, David E.  
REGISTRATION NUMBER: 22,592  
REFERENCE/DOCKET NUMBER: CMCC-614  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781) 861-6240  
TELEFAX: (781) 861-9540  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 22 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1  
OTHER INFORMATION: /label= modified aa  
OTHER INFORMATION: /note= "N-Acetyl-Threonine"  
US-09-474-743-7

Query Match 10.4%; Score 101; DB 4; Length 22;  
Best Local Similarity 100.0%; Pred. No. 8.4e-06;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 45 TFRAPLSSRLQDLXSYIVRAD 65

Db 1 TFRAPLSSRLQDLXSYIVRAD 21

RESULT 15  
US-08-740-168A-1  
Sequence 1, Application US/08740168A  
Patent No. 5854205  
GENERAL INFORMATION:  
APPLICANT: O'Reilly, Michael  
APPLICANT: Folkman, M. Judah  
TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions  
TITLE OF INVENTION: and Methods  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/740,168A  
FILING DATE: 22-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0223  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: Internal.  
ORIGINAL SOURCE:  
ORGANISM: Murine  
TISSUE TYPE: Collagen  
US-08-740-168A-1

Query Match 9.9%; Score 96; DB 2; Length 20;  
Best Local Similarity 85.0%; Pred. No. 3e-05;  
Matches 17; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HSHRDFQPVHLVALNSPLS 20  
1 TFRAPLSSRLQDLXSYIVRAD 21

Search completed: August 12, 2002, 10:13:26  
Job time: 138 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 12, 2002, 10:10:43 ; Search time 30.23 Seconds  
(without alignments)  
672.395 Million cell updates/sec

Title: US-10-080-797-1  
Perfect score: 967  
Sequence: 1 HSHRDFQPVYHLVALNSPLS.....SCHHAYIVLCIENSFWTASK 183

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                  |
|------------|-------|-------------|--------|-------|------------------------------|
| 1          | 967   | 100.0       | 183    | 20    | AAV08693 Human endostatin p  |
| 2          | 967   | 100.0       | 183    | 20    | AAV02113 SPO ID 76 of WO991  |
| 3          | 967   | 100.0       | 183    | 21    | AAAB30493 Amino acid sequenc |
| 4          | 967   | 100.0       | 183    | 21    | AAAB16451 Human endostatin p |
| 5          | 967   | 100.0       | 183    | 21    | AAAY90771 Human angiotensin  |
| 6          | 967   | 100.0       | 183    | 21    | AAAY70252 Human angiotensin  |
| 7          | 967   | 100.0       | 183    | 22    | AAU00896 Human Endostatin(T  |
| 8          | 967   | 100.0       | 183    | 22    | AAAB49379 Human endostatin S |
| 9          | 967   | 100.0       | 216    | 21    | AAAB30495 Amino acid sequenc |
| 10         | 967   | 100.0       | 684    | 18    | AAW26327 Human alpha-1 coll  |
| 11         | 967   | 100.0       | 684    | 20    | AAV25113 Human alpha1 (XVII  |

|    |     |       |      |    |                              |
|----|-----|-------|------|----|------------------------------|
| 12 | 967 | 100.0 | 1301 | 20 | AAW22296 Human alpha-1 (XVI  |
| 13 | 967 | 100.0 | 1336 | 20 | AAV08694 Human collagen 18   |
| 14 | 963 | 99.6  | 183  | 22 | AAAB49810 Human endostatin p |
| 15 | 962 | 99.5  | 182  | 21 | AAAB28399 Human endostatin.  |
| 16 | 962 | 99.5  | 182  | 21 | AAV94323 Human endostatin p  |
| 17 | 962 | 99.5  | 182  | 21 | AAV59622 Human endostatin p  |
| 18 | 962 | 99.5  | 182  | 21 | AAV59622 Human endostatin p  |
| 19 | 962 | 99.5  | 182  | 21 | AAU00897 Human Endostatin(T  |
| 20 | 960 | 99.3  | 195  | 21 | AAW90874 Human HMW endostat  |
| 21 | 958 | 99.1  | 184  | 22 | AAV78717 Human vascular end  |
| 22 | 956 | 98.9  | 181  | 22 | AAU00898 Human Endostatin(T  |
| 23 | 954 | 98.7  | 180  | 22 | AAU00899 Human Endostatin(T  |
| 24 | 954 | 98.7  | 193  | 21 | AAW90877 Human HMW endostat  |
| 25 | 942 | 97.4  | 179  | 22 | AAU00901 Human Endostatin(T  |
| 26 | 937 | 96.9  | 178  | 21 | AAV94324 Alternate human en  |
| 27 | 937 | 96.9  | 178  | 22 | AAU00900 Human Endostatin(T  |
| 28 | 846 | 87.5  | 684  | 20 | AAV25114 Mouse alpha1 (XVII  |
| 29 | 845 | 87.4  | 184  | 21 | AAV70265 Canine angiotensin  |
| 30 | 840 | 86.9  | 191  | 21 | AAAB49380 Murine endostatin  |
| 31 | 840 | 86.9  | 191  | 21 | AAAB28398 Murine endostatin  |
| 32 | 837 | 86.6  | 184  | 20 | AAV18409 Endostatin protein  |
| 33 | 837 | 86.6  | 184  | 20 | AAV08689 Murine endostatin   |
| 34 | 837 | 86.6  | 184  | 21 | AAV70258 Murine angiotensin  |
| 35 | 837 | 86.6  | 207  | 22 | AAE02031 Murine endostatin   |
| 36 | 837 | 86.6  | 207  | 22 | AAAB71930 Murine endostatin  |
| 37 | 837 | 86.6  | 218  | 20 | AAV08691 Murine gene therap  |
| 38 | 837 | 86.6  | 580  | 20 | AAV08692 Murine gene therap  |
| 39 | 837 | 86.6  | 1288 | 18 | AAW26328 Mouse alpha-1 coll  |
| 40 | 837 | 86.6  | 1288 | 20 | AAW22297 Mouse alpha-1 (XVI  |
| 41 | 796 | 82.3  | 185  | 20 | AAV06197 Anti-angiogenic en  |
| 42 | 771 | 79.7  | 184  | 22 | AAAB49381 Chicken endostatin |
| 43 | 544 | 56.3  | 180  | 22 | AAAB49383 Murine endostatin  |
| 44 | 529 | 54.7  | 176  | 21 | AAW90875 Human HMW endostat  |
| 45 | 529 | 54.7  | 180  | 22 | AAAB49382 Human endostatin S |

ALIGNMENTS

|          |  |
|----------|--|
| RESULT 1 |  |
| AAV08693 | AAV08693 standard; Protein; 183 AA.                                    |
| XX       | XX   |
| AC       | AAV08693;  |
| XX       | XX   |
| DT       | 10-AUG-1999 (first entry)  |
| XX       | XX   |
| DE       | Human endostatin protein fragment.                                     |
| KW       | Plasminogen; human; angiotensin; endostatin; gene therapy; vector;     |
| KW       | anti-angiogenic; attenuation; cytosolic; anti-diabetic; ophthalmology; |
| KW       | tumour growth; solid tumour; diabetic retinopathy; retina.             |
| OS       | Homo sapiens.  |
| XX       | XX   |
| PN       | WO9926480-A1.  |
| XX       | XX   |
| PD       | 03-JUN-1999.   |
| XX       | XX   |
| PF       | 20-NOV-1998; 98WO-US24950.   |
| XX       | XX   |
| PR       | 20-NOV-1997; 97US-0975424.   |
| PA       | (GENE-) GENETIX PHARM INC.   |
| PA       | (MAST ) MASSACHUSETTS INST TECHNOLOGY.                                 |
| XX       | XX   |
| PI       | Bachelot T, Leboulch P, Pawluc R;                                      |
| XX       | XX   |
| DR       | WPI: 1999-357696/30.   |
| XX       | XX   |
| PT       | N-PSDB: AAX77719.  |
| XX       | XX   |
| PT       | Anti-angiogenic gene therapy vectors                                   |
| XX       | XX   |

PS Disclosure; Page 74-75; 83pp; English.

CC This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine endostatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumor growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells ex vivo and then administered to the patient.

XX Sequence 183 AA;

Query Match 100.0%; Score 967; DB 20; Length 183;  
Best Local Similarity 100.0%; Pred. No. 9.8e-110;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGMRIGADFCQOARAVGLAGTFRATLSSRLQDLYSI 60  
Db 1 hshrdqpvhlvalnspisgmgirgadtgcqgaravglagtlratlssrlqdllysi 60  
QY 61 VRADRAAVPIVNLKDELFPSEWALFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSW 120  
Db 61 vrradraavpiylvnlkdelifpswealfsgseglpkpgarifsfdgkdvlrhptwpksw 120  
QY 121 HSDPNGRRLTESYCEFTWRREAPSATGQASSLLGRLGQSAASHAHAYVLCIENSFMT 180  
Db 121 hgsdpngrlrtesycefwtreapscatqgaasllgrrllgqsaaschahayivlcienfmc 180  
QY 181 ASK 183  
Db 181 ask 183

RESULT 2

AAV02113  
ID AAV02113 standard; protein; 183 AA.

XX AAV02113;

DT 16-JUL-1999 (first entry)

XX SEQ ID 76 of WO9916889.

XX Angiostatin; endostatin; interferon; thrombospondin;  
KW interferon-inducible protein; platelet factor 4; anti-angiogenic;  
KW anti-tumor; multifunctional protein; angiogenic-mediated disease;  
KW cancer; diabetic retinopathy; macular degeneration; arthritis;  
KW tumor cell production.

XX Homo sapiens.

OS Homo sapiens.

PN WO9916889-A1.

PD 08-APR-1999.

PF 30-SEP-1998; 98WO-US20464.

PR 01-OCT-1997; 97US-0060609.

XX (SEAR ) SEARLE & CO G D.

PI Bolanowski MA, Caparon MH, Casperson GF, Gregory SA;

PI Klein BK, Mckearn JP;

XX WPI; 1999-255098/21.

XX New multifunctional proteins useful for treating angiogenic-mediated

PT diseases

XX Disclosure; Page 106-107; 121pp; English.

CC The specification describes multifunctional proteins which comprise  
CC combinations of angiostatin, endostatin, interferon, thrombospondin,  
CC interferon-inducible protein and platelet factor 4, and have  
CC anti-angiogenic and/or anti-tumor activity. The multifunctional protein  
CC may exhibit useful properties such as having similar or greater  
CC biological activity when compared to a single factor or by having  
CC improved half-life or decreased adverse side effects, or a combination  
CC of these properties. The proteins can be used for treating an  
CC angiogenic-mediated disease, e.g. cancer, diabetic retinopathy, macular  
CC degeneration, or arthritis. They can also be used for inhibiting the  
CC production of tumor cells (characteristic of lung, breast, ovarian,  
CC prostate, pancreatic, gastric, colon, renal, bladder cancers; melanoma,  
CC hepatoma, sarcoma and lymphoma) in a patient and for inhibiting tumor  
CC growth. The present sequence is used in the course of the invention.

XX Sequence 183 AA;

Query Match 100.0%; Score 967; DB 20; Length 183;  
Best Local Similarity 100.0%; Pred. No. 9.8e-110;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDQPVHLVALNSPLSGMRIGADFCQOARAVGLAGTFRATLSSRLQDLYSI 60  
Db 1 hshrdqpvhlvalnspisgmgirgadtgcqgaravglagtlratlssrlqdllysi 60  
QY 61 VRADRAAVPIVNLKDELFPSEWALFSGSEGLPKPGARIFSPDGKDVLRHPTWPKSW 120  
Db 61 vrradraavpiylvnlkdelifpswealfsgseglpkpgarifsfdgkdvlrhptwpksw 120  
QY 121 HSDPNGRRLTESYCEFTWRREAPSATGQASSLLGRLGQSAASHAHAYVLCIENSFMT 180  
Db 121 hgsdpngrlrtesycefwtreapscatqgaasllgrrllgqsaaschahayivlcienfmc 180  
QY 181 ASK 183  
Db 181 ask 183

RESULT 3

AAB30493  
ID AAB30493 standard; protein; 183 AA.

XX AAB30493;

DT 06-MAR-2001 (first entry)

XX Amino acid sequence of human endostatin encoded by plasmid pMALCH#15.

KW Streptomyces sp. strain C5; SnPA; S. venezuelae; alpha-amylase;

KW endostatin; cancer; tumour growth; angiogenesis.

XX Homo sapiens.

OS Homo sapiens.

PN WO200060945-A1.

PD 19-OCT-2000.

PF 12-APR-2000; 2000WO-US09747.

PR 13-APR-1999; 99US-0129084.

XX (MERI ) MERCK & CO INC.

PI Desanti CL, Strohl WR;

XX WPI; 2000-686970/67.

XX N-PSDB; AAC62023.

PT Preparation of soluble recombinant endostatin involves transforming  
 PT Streptomyces host with expression vector comprising nucleotide  
 PT sequence encoding endostatin operably linked to linker and leader  
 PT peptide -

XX Example 1; Fig 6; 57pp; English.

XX The present sequence represents human endostatin. The protein is  
 CC expressed in Streptomyces. Leader sequences of Streptomyces sp. strain  
 CC C5 Sppa and S. venezuelae alpha-amylase proteins are linked to the  
 CC N-terminal of endostatin. This ensures that endostatin protein is  
 CC produced as a secreted, soluble protein which needs no refolding, is  
 CC stable in the fermentation broth and is produced in large quantities.  
 CC The method is used for preparing soluble recombinant human, murine or  
 CC primate endostatin, which is useful in the treatment of cancer.  
 CC Inhibition of tumour growth, inhibition of angiogenesis, isolation of  
 CC receptors for endostatin and for identification of anti-angiogenic  
 CC compounds in assays. The endostatin protein is produced as a secreted,  
 CC soluble protein which needs no refolding, is stable in the fermentation  
 CC broth and is produced in large quantities. Streptomyces are amenable  
 CC for cultivation in large fermentations allowing for large quantities of  
 CC soluble endostatin to be produced.

XX Sequence 183 AA:

Query Match 100.0%; Score 967; DB 21; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-110;  
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDFQVPLHLVALNSPLSGMGRGIRGADFOQFOARAVGLAGTFRAFLSSRLQDLYSI 60  
 Db 1 hshrdfqpvlhlvalnspisgmgirgadfcfgqaravglagtfraflssrlqdlysi 60  
 QY 61 VRRADRAAVPIVNLKDELLPSPWEALFSGSGEGLKPGARIFSDGKDYLRHPTWQKSVW 120  
 Db 61 vrradraavpivnlkdelllpsswealfsgsegplkpgarifsfdgkdvlrhptwpqksvw 120  
 QY 121 HGSDPNGRRLTESYCEWTRTEAPSATGQASSILGGRLLGQSAASHAHAYIVLCIENSFMT 180  
 Db 121 hgspnrgrrltesycewtrteapseatgqassilggrlllgqsaashahayivlcienfimt 180  
 QY 181 ASK 183  
 Db 181 ask 183

RESULT 4

AAB16451  
 ID AAB16451 standard; Protein; 183 AA.

XX AAB16451.

DT 27-OCT-2000 (first entry)

XX Human endostatin protein sequence.

XX Angiogenesis-inhibiting protein receptor; angiogenesis; anglostatin;  
 KW endostatin; plasminogen; laminin; treatment; wound healing; solid tumour;  
 KW psoriasis; scleroderma; myocardial angiogenesis; Crohn's disease;  
 KW cerebral collateral; arteriovenous malformation; rubecosis; cancer;  
 KW diabetic retinopathy; arthritis; wound healing; peptic ulcer;  
 KW Helicobacter related disease; fracture; cat scratch fever.

XX Homo sapiens.

XX WO2000032631-A2.

XX 08-JUN-2000.

XX 06-DEC-1999; 99WO-US28897.

XX 04-DEC-1998; 98US-0206059.

XX (ENTR-) ENTREMED INC.  
 PA Macdonald NJ, Sim KL;  
 PI WPI; 2000-412290/35.

XX New angiogenesis-inhibiting protein receptors, useful in methods for  
 PT treating diseases and processes that are mediated by angiogenesis, such  
 PT as solid tumours, psoriasis, scleroderma and myocardial angiogenesis -  
 XX Disclosure; Figure 3; 100pp; English.

XX This invention relates to angiogenesis-inhibiting protein receptors, and  
 CC the DNA sequences encoding them. Angiogenesis is the generation of new  
 CC blood vessels into a tissue, and normally occurs in wound healing,  
 CC foetal and embryonal development and the formation of the corpus luteum,  
 CC endometrium and placenta. Anglostatin is a protein (see AAB16450 and  
 CC AA68202) involved in angiogenesis, and has an amino acid sequence  
 CC similar to that of a plasminogen fragment (see murine plasminogen  
 CC AAB16490). Anglostatin has the ability to inhibit angiogenesis.  
 CC Endostatin is also an angiogenesis inhibiting protein (see AAB16451 and  
 CC AA68203). Sequences AA68242 and AAB16522 represent coding and protein  
 CC sequences of human laminin. Laminin is an anglostatin binding protein,  
 CC and some of the peptides of the invention share homology with regions of  
 CC laminin. Peptides AAB16452-B16521 (excluding AAB16490) are the  
 CC angiogenesis-inhibiting protein receptor fragments of the invention. The  
 CC peptides bind either anglostatin or endostatin and can be used in methods  
 CC for treating diseases and processes that are mediated by angiogenesis,  
 CC such as solid tumours, psoriasis, scleroderma, myocardial angiogenesis,  
 CC Crohn's disease, cerebral collaterals, arteriovenous malformations,  
 CC rubecosis, diabetic retinopathy, arthritis, wound healing, peptic ulcers,  
 CC Helicobacter related diseases, fractures, placentaion and cat scratch  
 CC fever. They are useful for the detection and prognosis of cancer. DNA  
 CC sequences A628204-A628241 encode the peptides of the invention.

XX Sequence 183 AA:

Query Match 100.0%; Score 967; DB 21; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 9.8e-110;  
 Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDFQVPLHLVALNSPLSGMGRGIRGADFOQFOARAVGLAGTFRAFLSSRLQDLYSI 60  
 Db 1 hshrdfqpvlhlvalnspisgmgirgadfcfgqaravglagtfraflssrlqdlysi 60  
 QY 61 VRRADRAAVPIVNLKDELLPSPWEALFSGSGEGLKPGARIFSDGKDYLRHPTWQKSVW 120  
 Db 61 vrradraavpivnlkdelllpsswealfsgsegplkpgarifsfdgkdvlrhptwpqksvw 120  
 QY 121 HGSDPNGRRLTESYCEWTRTEAPSATGQASSILGGRLLGQSAASHAHAYIVLCIENSFMT 180  
 Db 121 hgspnrgrrltesycewtrteapseatgqassilggrlllgqsaashahayivlcienfimt 180  
 QY 181 ASK 183  
 Db 181 ask 183

RESULT 5

AA90771  
 ID AA90771 standard; Protein; 183 AA.

XX AA90771;

XX 22-AUG-2000 (first entry)

XX Human angiogenesis inhibiting factor 1 protein.

XX Human; angiogenesis inhibiting factor 1; IAF-1; tumour; antibody;  
 KW abnormal vessel disease.

OS Homo sapiens.  
XX  
XX CN1244536-A.  
XX  
PD 16-FEB-2000.  
XX  
PF 10-AUG-1998: 98CN-0117150.  
XX  
PR 10-AUG-1998: 98CN-0117150.  
XX  
PA (ONCO-) INST ONCOLOGY UNDER TUMOR HOSPITAL CHINE.  
XX  
PI Yang Z, Guo W;  
XX  
XX WPI: 2000-388168/34.  
DR N-PSDB: AAA29884.  
XX  
PT Angiogenesis inhibiting factor 1 and its derivative useful for treating  
PT tumors -  
XX  
PS Claim 1; Fig 5; 41pp; Chinese.  
XX  
CC The present sequence represents an angiogenesis inhibiting factor (1),  
CC designated IAF-1. The present invention also describes: (1) preparation  
CC of (1) and its derivative; (2) an IAF binding acceptor and its  
CC preparation; and (3) an IAF antibody. (1) is useful for preparing new  
CC biological preparations for effectively treating various tumours and  
CC abnormal-vessel diseases. The IAF antibody is preferably a polyclonal  
CC antibody, mosaic antibody, single stranded antibody and human originated  
CC antibody.  
XX  
SQ Sequence 183 AA:  
  
Query Match 100.0%; Score 967; DB 21; Length 183;  
Best Local Similarity 100.0%; Pred. No. 9.8e-110;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HSHRDFOPLHLVALNSPLSGMKGIRGADFOCFQOARAVGLGTFRAFLSSRLQDLYST 60  
DB 1 hshrdfgpyvlhlvalnspisgmgirgadfgcfqaravglagtrfafsrlqdllyst 60  
QY 61 VRRADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSGDKVLRHPMPQKSW 120  
DB 61 vrradraavpiivnlkdelfpsswealfsgseglpkparifsfdkvdlrhpmpqksw 120  
QY 121 HGSDPNGRRLTESYCEWTWRTEAPSATGQASSLLGRLGSAASHHAYIVLCIENSFMT 180  
DB 121 hgsdpngrrltesyccetwrteapsatgqassllgrrlggsaaschhayivlcienfnt 180  
QY 181 ASK 183  
DB 181 ask 183  
  
RESULT 6  
ID AAY70252 standard; Protein: 183 AA.  
XX  
AC AAY70252;  
XX  
DT 06-JUN-2000 (first entry)  
XX  
DE Human angiogenesis inhibitor, endostatin.  
XX  
KW Human: immunoglobulin gamma Fc fragment; endostatin; immunofusion;  
KW angiogenesis; inhibitor; cytoskeletal; antirheumatoid; antirheumatic;  
KW antiproliferative; antidiabetic; ophthalmological; immunosuppressant;  
KW vasculitic; vulnery; treatment; antiarteriosclerosis; tumour;  
KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
KW ocular angiogenesis; diabetic retinopathy; macular degeneration;  
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
KW wound granulation; keloid scar; gene therapy.

XX  
OS Homo sapiens.  
XX  
XX WO200011033-A2.  
XX  
PD 02-MAR-2000.  
XX  
PF 25-AUG-1999: 99WO-US19329.  
XX  
PR 25-AUG-1998: 98US-0097883.  
XX  
PA (LEXI-) LEXINGEN PHARM CORP.  
XX  
PI Lo R, Li Y, Gillies SD;  
XX  
XX WPI: 2000-237616/20.  
DR N-PSDB: AA251291.  
XX  
PT Novel fusion protein of angiotensin or endostatin and an immunoglobulin  
PT Fc region, useful for treating conditions mediated by angiogenesis,  
PT such as rheumatoid arthritis, tumors and macular degeneration -  
XX  
PS Example 1; Pages 41-42; 68pp; English.  
XX  
CC The patent discloses a DNA molecule encoding a fusion protein comprising  
CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis  
CC inhibitor selected from angiotensin, endostatin, a plasminogen fragment  
CC having angiotensin activity, a collagen XVIII fragment having endostatin  
CC activity, or combinations of them. The fusion protein (immunofusion) is  
CC used to inhibit angiogenesis and to treat diseases or conditions mediated  
CC by angiogenesis. Conditions that may be treated include solid tumours,  
CC blood born tumours, tumour metastasis, benign tumours including  
CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic  
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases  
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular  
CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental  
CC fibroplasia, rubeosis and Osler-Webber syndrome, myocardial angiogenesis,  
CC plaque neovascularisation, telangiectasia, haemophilic joints,  
CC angiodysplasia, wound granulation, and excessive or abnormal stimulation of  
CC endothelial cells, intestinal cells, atherosclerosis, scleroderma and  
CC hypertrophic scars, i.e. keloid scars. The DNA constructs may be used  
CC in gene therapy. The present sequence is a human endostatin used in the  
CC construction of immunofusion containing human immunoglobulin gamma  
CC (19c) Fc fragment.  
XX  
SQ Sequence 183 AA:  
  
Query Match 100.0%; Score 967; DB 21; Length 183;  
Best Local Similarity 100.0%; Pred. No. 9.8e-110;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HSHRDFOPLHLVALNSPLSGMKGIRGADFOCFQOARAVGLGTFRAFLSSRLQDLYST 60  
DB 1 hshrdfgpyvlhlvalnspisgmgirgadfgcfqaravglagtrfafsrlqdllyst 60  
QY 61 VRRADRAAVPIVNLKDELFPSSWEALFSGSEGPLKPGARIFSGDKVLRHPMPQKSW 120  
DB 61 vrradraavpiivnlkdelfpsswealfsgseglpkparifsfdkvdlrhpmpqksw 120  
QY 121 HGSDPNGRRLTESYCEWTWRTEAPSATGQASSLLGRLGSAASHHAYIVLCIENSFMT 180  
DB 121 hgsdpngrrltesyccetwrteapsatgqassllgrrlggsaaschhayivlcienfnt 180  
QY 181 ASK 183  
DB 181 ask 183  
  
RESULT 7  
ID AAU00896 standard; Protein: 183 AA.  
XX

AAU00896;  
04-JUL-2001 (first entry)  
Human Endostatin(TM) protein.  
Human; Endostatin(TM); angiogenesis mediated disease; solid tumours;  
blood borne tumour; leukaemia; tumour metastasis; benign tumour;  
hemangioma; acoustic neuroma; neurofibroma; trachoma; rubecosis;  
pyogenic granuloma; rheumatoid arthritis; psoriasis; colon cancer;  
ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
retinopathy of prematurity; macular corneal rejection;  
neovascular glaucoma; retrolental fibroplasia; Osler-Weber Syndrome;  
myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
hemophiliac joint; angiofibroma; wound granulation.  
Homo sapiens.  
WO200119989-A2.  
22-MAR-2001.  
14-SEP-2000; 2000WO-US25166.  
14-SEP-1999; 99US-0153698.  
(ENTR-) ENTREMED INC.  
Liang H, Sim KL, Chang-Murad A, Zhou X, Madsen J, Boerner RJ;  
Bermejo LL, Mistry FR, Shepard SR, Schrimsher JL;  
WPI; 2001-244802/25.  
N-PSDB; AAC00867.  
Producing Endostatin protein for treating angiogenesis mediated  
diseases such as solid tumours, comprises recombinantly producing the  
protein using an expression system, and recovering and purifying the  
protein -  
Claim 5; Page 29; 67pp; English.  
The sequence represents Human Endostatin(TM). The new method of the  
invention is useful for producing, recovering and purifying Endostatin  
(TM) from biological sources, such as biological fluids, tissues, cells,  
culture media, and fermentation media. Endostatin(TM) is useful for  
treating angiogenesis mediated diseases such as solid tumours, blood  
borne tumours, leukaemias, tumour metastases, benign tumours, e.g.  
hemangioma, acoustic neuromas, neurofibromas, trachomas, and pyogenic  
granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases,  
e.g., diabetic retinopathy, retinopathy of prematurity, macular  
degeneration, corneal graft rejection, neovascular glaucoma, colon  
cancer, retrolental fibroplasia, rubecosis, Osler-Weber Syndrome,  
myocardial angiogenesis, plaque neovascularisation, telangiectasia,  
hemophiliac joints, angiofibroma, and wound granulation. Endostatin(TM)  
is also useful for treating disease of excessive or abnormal stimulation  
of endothelial cells such as intestinal adhesions, atherosclerosis,  
scleroderma and hypertrophic scars. Higher yields of more purified, and  
biologically active Endostatin(TM) are obtained by the new method.  
Endostatin(TM) can be stored in buffers for extended periods of time, and  
also subjected to lyophilisation, while preserving biological activity.  
Centrifugation of broth from fermentation steps in production is avoided,  
preventing unwanted potential cellular lysis and contamination with  
additional proteins, pigments, enzymes and other cellular chemicals and  
debris.  
Sequence 183 AA;  
SQ

Query Match 100.0%; Score 967; DB 22; Length 183;  
Best Local Similarity 100.0%; Pred. No. 9,8e-110;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDFPVHLVALNSPLSGMGRGADFGQFQOARAVGLACTFRAFLSSRLQDIYSI 60

|||||  
Db 1 hshrdfpvhlvalnspisgmgirgadfgcfqgaravglaqfiraflssrlqdiysi 60  
QY 61 VRRADRAAVPIVNLKDELLFPSEALFSGSGPLKPGARIRFSFGKDVLRHPTPQKSVW 120  
|||||  
Db 61 vrradraavpivnlkdelllfswaelfsgsgplkpgarirfsfdgkdvlrhptwpqksvw 120  
QY 121 HGSDPNGRRLNESCEMTRETPASATGQASSLIGRILGQSAASCHHAYITLCTENSPMT 180  
|||||  
Db 121 hgsdpngrlrlescemtretpasatgqassllgrrllgqaaschnayivlctensmtc 180  
QY 181 ASK 183  
|||  
Db 181 ask 183  
RESULT 8  
AAB49379  
ID AAB49379 standard; Protein; 183 AA.  
XX  
AC AAB49379;  
XX  
DT 02-MAR-2001 (first entry)  
XX  
DE Human endostatin SEQ ID NO: 2.  
XX  
KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;  
KW cancer; inflammation; angiogenesis-dependent disease.  
OS Homo sapiens.  
XX  
PN WO200067771-A1.  
XX  
PD 16-NOV-2000.  
XX  
PF 02-MAY-2000; 2000WO-US12063.  
XX  
PR 06-MAY-1999; 99US-0132907.  
PR 14-JUL-1999; 99US-0353333.  
XX  
PA (BURN-) BURNHAM INST.  
XX  
PI Vuori K;  
XX  
DR WPI; 2001-040937/05.  
DR N-PSDB; AAC88289.  
XX  
PT Endostatin peptide comprising at least four endostatin amino acid  
PT residues are e.g. angiogenesis inhibitors for treating cancer and  
PT diabetic retinopathy -  
XX  
PS Disclosure; Fig 1; 146pp; English.  
XX  
XX The present invention provides endostatin peptides which can be used in  
CC the modulation of angiogenesis. This is useful in the treatment of  
CC cancers, inflammation, rheumatoid arthritis, chronic articular  
CC rheumatism, psoriasis, disorders associated with inopportune invasion of  
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy  
CC of prematurity, macular degeneration, corneal graft rejection,  
CC retrolental fibroplasia, rubecosis, capillary proliferation in  
CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent  
CC diseases include Osler-Weber syndrome, myocardial angiogenesis, plaque  
CC neovascularisation, telangiectasia, hemophiliac joints and wound  
CC granulation. In addition, the peptides can be used as birth control  
CC agents.  
XX  
SQ Sequence 183 AA;  
QY

Query Match 100.0%; Score 967; DB 22; Length 183;  
Best Local Similarity 100.0%; Pred. No. 9,8e-110;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDFQPVHLVALNSPLSGMRGIRGADFOCFQOARAVGLAGTFRAPLSSRLDLYSI 60  
DB 1 hshrdfgpvhlvalnspisgmrgirgadfgcfqgaravglagtfraflsrlqdlysi 60  
QY 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGLKPGARIFSDGKDVLRHPMPQKSW 120  
DB 61 vrradraavpivnlkdel lfpsewal fsgseglkpgar ifsdgkdv lrhptwpqksv 120  
QY 121 HGSDPNGRRLTESYCETWRTAPSATGOASSLLGGRLLGOSASCHHAYIVLCIENSFMT 180  
DB 121 hgsdpngrrltesycetwrtap satgoassllggrllggsaaschhayivlcien sfmt 180  
QY 181 ASK 183  
DB 181 ask 183  
RESULT 9  
AAB30495  
ID AAB30495 standard; protein; 216 AA.  
XX  
AC AAB30495;  
XX  
DT 06-MAR-2001 (first entry)  
XX  
DE Amino acid sequence of vaa-endostatin fusion protein in pANT3052.  
XX  
KM Streptomyces sp. strain C5; SnPA; S. venezuelae; alpha-amy-lase;  
XX endostatin; cancer; tumour growth; angiogenesis.  
OS Synthetic.  
OS Streptomyces sp.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Peptide 1..28 /note= "vaa signal sequence"  
FT Protein 29..216 /note= "endostatin"  
XX  
PN WO200060945-A1.  
XX  
PD 19-OCT-2000.  
XX  
PF 12-APR-2000; 2000MO-US09747.  
XX  
PR 13-APR-1999; 99US-0129084.  
XX  
PA (MERI ) MERCK & CO INC.  
XX  
PI Desanti CL, Strohl WR;  
XX  
DR WPI: 2000-686970/67.  
DR N-PSDB; AAC62025.  
XX  
PT Preparation of soluble recombinant endostatin involves transforming  
PT Streptomyces host with expression vector comprising nucleotide  
PT sequence encoding endostatin operably linked to linker and leader  
PT peptide  
XX  
PS Example 1; Fig 10A-B; 57pp; English.  
XX  
CC The present sequence represents a fusion protein of vaa and endostatin.  
CC The specification describes a method for the production of soluble,  
CC recombinant human endostatin in Streptomyces. Leader sequences of  
CC Streptomyces sp. strain C5 SnPA and S. venezuelae alpha-amy-lase proteins  
CC are linked to the N-terminal of endostatin. This ensures that endostatin  
CC protein is produced as a secreted, soluble protein which needs no  
CC refolding, is stable in the fermentation broth and is produced in large  
CC quantities. The method is used for preparing soluble recombinant human,  
CC murine or primate endostatin, which is useful in the treatment of cancer,  
CC inhibition of tumour growth, inhibition of angiogenesis, isolation of  
CC receptors for endostatin and for identification of anti-angiogenic

CC compounds in assays. The endostatin protein is produced as a secreted,  
CC soluble protein which needs no refolding, is stable in the fermentation  
CC broth and is produced in large quantities. Streptomyces are amenable  
CC for cultivation in large fermentations allowing for large quantities of  
CC soluble endostatin to be produced.  
XX  
SQ Sequence 216 AA:  
QY 1 HSHRDFQPVHLVALNSPLSGMRGIRGADFOCFQOARAVGLAGTFRAPLSSRLDLYSI 60  
DB 34 hshrdfgpvhlvalnspisgmrgirgadfgcfqgaravglagtfraflsrlqdlysi 93  
QY 61 VRRADRAAVPIVNLKDELLFPSSWEALFSGSEGLKPGARIFSDGKDVLRHPMPQKSW 120  
DB 94 vrradraavpivnlkdel lfpsewal fsgseglkpgar ifsdgkdv lrhptwpqksv 153  
QY 121 HGSDPNGRRLTESYCETWRTAPSATGOASSLLGGRLLGOSASCHHAYIVLCIENSFMT 180  
DB 154 hgsdpngrrltesycetwrtap satgoassllggrllggsaaschhayivlcien sfmt 213  
QY 181 ASK 183  
DB 214 ask 216  
RESULT 10  
AAM26327  
ID AAM26327 standard; protein; 684 AA.  
XX  
AC AAM26327;  
XX  
DT 19-NOV-1997 (first entry)  
XX  
DE Human alpha-1 collagen (XVIII).  
XX  
KM Alpha-1 collagen; type XVIII collagen; cartilage degeneration.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..6 /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 7..12 /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 13..18 /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 19..24 /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 25..30 /label= GYXG'Y'\_\_motif  
FT /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 31..36 /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 37..42 /label= GYXG'Y'\_\_motif  
FT /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 48..53 /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 54..59 /label= GYXG'Y'\_\_motif  
FT /label= GYXG'Y'\_\_motif  
FT /note= "Claim 1"  
FT 74..79  
FT Peptide

[illegible]

XX Novel human type alpha-1 (XVIII) collagen is characterised by  
CC 10 triple helical domains containing the GXYGX'Y' motif (where X,  
CC Y, X' and Y' represent any amino acid), the helical domains being  
CC separated and flanked by non-triple helical regions which may  
CC provide flexibility. Alpha-1 collagen is expressed in multiple  
CC tissues, especially liver, lung and kidney. A claimed plasmid  
CC comprising alpha-1 collagen nucleic acid (see AAT84484) and an  
CC expression control sequence can be used to express recombinant  
CC collagen in prokaryotic or eukaryotic (especially mammalian) host  
CC cells. The alpha-1 collagen may be used to treat a patient  
CC suffering from a disease associated with cartilage degradation, and  
CC for supplementing collagen. It can also be used as a connective

Query Match 100.0%; Score 967; DB 18; Length 684;

Best Local Similarity 100.0%; Pred. No. 7e-109; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDPOVHLVALNSPLSGMRGTRGADPQCARAVGLACTPRAFLSSRLQDLYSI 60  
DB 502 hshrdtpvhlvalnspisgmrgtrgadpfcqgaravglactpfaflssrlqdllysl 561  
QY 61 VRRADRAAVPIVNLKDELPPSEWALFSGSEGPLKPGARIFSPDGKDVLRHPMPQKSW 120  
DB 562 vrradraavpivnlkdelppsewaltsgeegplkpgarilfsdpgkdvlrhpcwpqksw 621  
QY 121 HGSDPNGRRLTESYCETWRTAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180  
DB 622 hgsdpgngrrltesycetwrtaprsatgqassllggrllggsaaschhayivlcienfnt 681  
QY 181 ASK 183  
DB 682 ask 684

## RESULT 11

AA25113  
ID AAY25113 standard; Protein; 684 AA.

AC AAY25113;

DT 25-AUG-1999 (first entry)

DE Human alpha1 (XVIII) collagen protein.

XX Alpha1(XVIII) collagen; mimetic; endostatin; atomic coordinate; library;  
KW anti-angiogenic; heparin binding domain; receptor binding domain; mimic;  
KW alpha-helix A domain; carbohydrate recognition domain; CRD domain;  
KW treatment; angiogenesis; tumour; human.

XX Homo sapiens.

OS WO9931616-A1.

PN 24-JUN-1999.

PD 16-DEC-1998; 98WO-US26783.

PR 16-DEC-1997; 97US-0069727.

XX (HARD ) HARVARD COLLEGE.

PI Hohenester E, Olsen BR, Sasaki T, Timpl R;

DR WPI; 1999-395243/33.

XX N-PSDB; AAX78379.

PT Identifying mimetics of mammalian endostatin

PS Disclosure; Fig 5A-C; 75pp; English.

CC This invention describes a novel method for identifying mimetics of  
CC mammalian endostatin. The method comprises identifying a compound

CC having atomic coordinates with non-trivial similarity to selected  
CC coordinates of atoms of a mammalian endostatin involves (a) providing  
CC a library of atomic coordinates of compounds in a library of candidate  
CC compounds, (b) comparing the library of atomic coordinates to the  
CC selected coordinates of a mammalian endostatin and (c) selecting from the  
CC library at least one candidate compound on the basis of selection  
CC criteria which include similarities between the atomic coordinates of the  
CC selected candidate compound and the atomic coordinates of the mammalian  
CC endostatin. The invention also describes the use of an anti-angiogenic  
CC fragment of endostatin comprising a domain selected from a heparin  
CC binding domain, a receptor binding domain, and exposed on alpha-helix A  
CC domain, and a carbohydrate recognition domain (CRD) domain. The methods  
CC can be used for designing and selecting undesired angiogenesis, e.g. tumours.  
CC This sequence represents human alpha1(XVIII) collagen which is used in  
CC the description of the method.

SO Sequence 684 AA;

Query Match 100.0%; Score 967; DB 20; Length 684;

Best Local Similarity 100.0%; Pred. No. 7e-109; Mismatches 0; Indels 0; Gaps 0;

QY 1 HSHRDPOVHLVALNSPLSGMRGTRGADPQCARAVGLACTPRAFLSSRLQDLYSI 60  
DB 502 hshrdtpvhlvalnspisgmrgtrgadpfcqgaravglactpfaflssrlqdllysl 561  
QY 61 VRRADRAAVPIVNLKDELPPSEWALFSGSEGPLKPGARIFSPDGKDVLRHPMPQKSW 120  
DB 562 vrradraavpivnlkdelppsewaltsgeegplkpgarilfsdpgkdvlrhpcwpqksw 621  
QY 121 HGSDPNGRRLTESYCETWRTAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180  
DB 622 hgsdpgngrrltesycetwrtaprsatgqassllggrllggsaaschhayivlcienfnt 681  
QY 181 ASK 183  
DB 682 ask 684

## RESULT 12

AAW92296  
ID AAW92296 standard; peptide; 1301 AA.

AC AAW92296;

DT 28-APR-1999 (first entry)

DE Human alpha-1 (XVIII) collagen chain common sequence HUI8(common)36.

XX Human; type XVIII collagen; liver disease; cirrhosis; detection;  
KW hepatocellular carcinoma; diagnosis.

XX Homo sapiens.

OS WO9856399-A1.

PN 17-DEC-1998.

PD 12-JUN-1998; 98WO-US12327.

PR 12-JUN-1997; 97US-0049369.

XX (FIFI-) ACAD FINLAND.

PA (FIBR-) FIBROGEN INC.  
PA (INRM ) INST NAT SANTE & RECH MEDICALE.

XX Clement B, Pihlajaniemi T, Rehn M;

DR WPI; 1999-070292/06.

PT Diagnosis and monitoring of liver disease by measuring collagen type



PT XVIII levels - with elevated levels indicative of disease,  
PT especially cirrhosis or hepatocellular carcinoma  
XX  
PS Example 6; Fig 8; 56pp; English.  
XX  
CC A method has been developed for the detecting liver disease. The method  
CC comprises: (a) reacting a patient sample with antibodies (Ab) specific  
CC for collagen type XVIII (Coll18); (b) measuring the amount of Ab-antigen  
CC complex (C) formed as indicator of the amount of Coll18 present; (c)  
CC similar analysis of a non-diseased control; and (d) comparing the  
CC amounts of Coll18 in the two samples to detect presence or progression of  
CC disease. Elevated levels of Coll18 are: (i) indicative of disease,  
CC specifically cirrhosis; and (ii) predictive of the prognosis of disease,  
CC specifically hepatocellular carcinoma (there is a relationship between  
CC Coll18 mRNA levels and tumour size and necrosis, and survival times are  
CC significantly higher in patients with higher Coll18 levels). The method  
CC provides non-invasive, early and accurate diagnosis of liver disease.  
CC The present sequence represents the sequence common to human alpha-1  
CC (XVIII) collagen chain from the present invention.  
XX  
SQ Sequence 1301 AA:  
  
Query Match 100.0%; Score 967; DB 20; Length 1301;  
Best Local Similarity 100.0%; Pred. No. 1.8e-108;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HSHRDQPVHLVNLNSPLSGMGIRGADFCQCARAVGLAGTFRAFLSSRIQDLYSI 60  
Db 1119 hshrdtqpvhlvnlhnlspisgmgirgadfcqgaravglagtfraflssriqdllysi 1178  
  
QY 61 VRRADRAAVPIVNLKDELFPSEWALFSGSEGPLKPGARIFSEFDGKDVLRHPTWPKSVW 120  
Db 1179 vrradraavpivlvnlkdelfpsewalfsgegplkpgarifsfqgkdvlrhptwpqksvw 1238  
  
QY 121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180  
Db 1239 hgsdpngrrltesyctewrtreapsatgqassllggrllggsaachayivlcienfnt 1298  
  
QY 181 ASK 183  
Db 1299 ask 1301  
  
RESULT 13  
AA08694  
ID AAY08694 standard; Protein: 1336 AA.  
XX  
AC AAY08694;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Human collagen 18 protein.  
XX  
KW Plasminogen; human; angiotatin; endostatin; gene therapy; vector;  
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;  
KW tumour growth; solid tumour; diabetic retinopathy; retina; collagen 18.  
XX  
OS Homo sapiens.  
XX  
PN WO926480-A1.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-US24950.  
XX  
PR 20-NOV-1997; 97US-0975424.  
XX  
PA (GENE-) GENETIX PHARM INC.  
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.  
PI Bachelot T, Leboulch P, Pawliuk RJ;  
XX

DR WPI; 1999-357696/30.  
DR N-PSDB; AAX77720.  
XX  
PT Anti-angiogenic gene therapy vectors  
XX  
PS Disclosure; Page 77-80; 83pp; English.  
XX  
CC This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine angiotatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells ex vivo and then administered to the patient.  
XX  
SQ Sequence 1336 AA:  
  
Query Match 100.0%; Score 967; DB 20; Length 1336;  
Best Local Similarity 100.0%; Pred. No. 1.9e-108;  
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 HSHRDQPVHLVNLNSPLSGMGIRGADFCQCARAVGLAGTFRAFLSSRIQDLYSI 60  
Db 1154 hshrdtqpvhlvnlhnlspisgmgirgadfcqgaravglagtfraflssriqdllysi 1213  
  
QY 61 VRRADRAAVPIVNLKDELFPSEWALFSGSEGPLKPGARIFSEFDGKDVLRHPTWPKSVW 120  
Db 1214 vrradraavpivlvnlkdelfpsewalfsgegplkpgarifsfqgkdvlrhptwpqksvw 1273  
  
QY 121 HGSDPNGRRLTESYCEWTREAPSATGQASSLLGGRLLGQSAASCHHAYIVLCIENSFMT 180  
Db 1274 hgsdpngrrltesyctewrtreapsatgqassllggrllggsaachayivlcienfnt 1333  
  
QY 181 ASK 183  
Db 1334 ask 1336  
  
RESULT 14  
AAB49810  
ID AAB49810 standard; Protein: 183 AA.  
XX  
AC AAB49810;  
XX  
DT 02-MAR-2001 (first entry)  
XX  
DE Human endostatin peptide fragment SEQ ID NO: 23.  
XX  
KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;  
KW cancer; inflammation; angiogenesis-dependent disease.  
XX  
OS Homo sapiens.  
XX  
PN WO200067771-A1.  
XX  
PD 16-NOV-2000.  
XX  
PF 02-MAY-2000; 2000WO-US12063.  
XX  
PR 06-MAY-1999; 99US-0132907.  
XX PR 14-JUL-1999; 99US-0353333.  
XX  
PA (BURN-) BURNHAM INST.  
XX  
PI Vuori K;  
XX  
DR WPI; 2001-040937/05.

